THE ROLE OF SEXUAL RECOMBINATION IN THE EVOLUTION OF THE PROTOZOAN PARASITE, *TOXOPLASMA GONDII*

by

Andrea Kennard

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

July 2018

Abstract

Toxoplasma gondii is an obligate intracellular protozoan parasite with a haploid genome. *Toxoplasma* has the capacity to infect any warm-blooded animal and infects between 30-80% of humans and 47% of wildlife surveyed globally. Although this parasite possesses an exceedingly flexible replication strategy, with both a prevalent asexual replication cycle in any intermediate host, and a fecund sexual cycle capable of producing in excess of 100 million infectious oocysts exclusively in its definitive felid host; the extent to which each replication method is utilized to transmit this generalist parasite in nature has long been under debate. Sexual recombination in *Toxoplasma* is an advantageous trait that allows the parasite to admix population genetic diversity to generate new biological potential, such as the ability to invade new hosts, evade host immunity, or cause outbreaks. However, despite its fecund sexual cycle, *Toxoplasma* has a genetically limited population structure that is dominated by just a few clonal lines that circulate predominantly in North America and Europe. Two theories exist to explain this clonal population structure, the first suggests that the clonal lines are highly adapted for oral transmission among intermediate hosts and are expanding exclusively asexually within this population of hosts, the second states that the clonal lines are particularly fit clones that are being expanded in cats by uniparental self-mating due to *Toxoplasma's* lack of mating types and its ability to undergo unisexual expansion via single clone infection in the definitive host. In support of the latter theory, fungal pathogens have been shown previously to cryptically expand certain clones when sexual recombination occurs between closely related strains. The recombinant progeny produced are typically indistinguishable from each other due to the use of limited, low-resolution markers that fail to capture the genetic differences between the two mating clones.

During the tenure of this thesis, studies have been published that chart *Toxoplasma* global population genetic diversity, but no study has examined intra-typic (*i.e.,* within a clonal lineage) population genetics at whole genome resolution. Work shown here interrogates the most prevalent clonal lineage characterized worldwide, Type II, and the most prevalent clonal lineage in North America, Type X. Work here aims to determine at whole genome resolution the intra-typic genetic relationship within two previously reported clonal lineages: Type II and Type X (also referred to as HG12), and whether strains within these two clonal groups isolated across different geographies and animal host species, are evolving via sexual recombination or genetic drift. Strains studied in this thesis were examined by limited sequencing at established *Toxoplasma* genotyping markers, as well as by whole genome sequencing, which has only recently come into use to interrogate genetic diversity and determine the extent to which recombination is occurring within the genome of this parasite. Type II strains were identified to have undergone limited mitotic drift within the lineage, and the polymorphism detected was generally correlated with the region of the strains' geographic isolation. However, distinct haploblocks of Type II sequences bearing the hallmarks of different geographic regions were identified in a subset of Type II isolates, and these admixtures suggested that intra-typic recombination (unisexual mating) was occurring between Type II strains of distinct geographies. Proof of unisexual mating between two highly homologous, but independent clones, was identified phylogenetically by an incongruence between nuclear and maternally inherited organellar genomes among these Type II strains. Importantly, genotyping using just ten low resolution PCR-RFLP markers generally predicted that the strain belonged to the Type II lineage, because relatively few admixtures between Type II and other genetic backgrounds were identified after increasing the resolution to WGS. Although at first glance this appears to support

a model whereby Type II strains bypass their sexual cycle, and are expanding largely asexually, closer examination of the haploblocks of geographically shared polymorphism identified a number of Type II strains that possessed different admixtures of these Type II, geographically derived haploblocks. This observation was parsimonious with recombination among clade-specific genotypes and supports expansion by unisexual mating.

The other clonotype supported by low-resolution PCR-RFLP analyses, and interrogated here, Type X, is the most prevalent genotype infecting wildlife in North America. This thesis performed an unprecedented, longitudinal study of isolates collected over a 7-year period from a single host species, Southern sea otters, in a localized geographic region. In contrast to the Type II lineage, outcrossing was especially pronounced among the Type X strains examined. In fact, Type X was determined to be a clade of sexually-related progeny from at least one sexual cross between a Type II ancestral strain and a previously unknown strain of mosaic ancestry, referred to as γ/δ. When infected in mice, Type X strains displayed a range of virulence phenotypes from highly virulent to avirulent, like a Type II strain. Because these natural isolates resembled recombinant progeny, the Type X strains were utilized as if they were true F1 progeny from a sexual cross and a QTL was performed to identify genes conferring mouse virulence. The analysis identified a novel virulence gene, ROP33. Evidence from whole genome sequencing demonstrated that while Type X strains resembled closely related progeny, mitotic drift had occurred within these strains prior to their infection in sea otters. However, the observed mitotic drift, in combination with the clear and limited number of crossover break points in Type X, supports a model in which only a limited number of crosses between Type II and a strain of distinct ancestry occurred to generate the Type X strains, but that the Type II strain that crossed with the γ/δ parent was divergent. Two mouse virulent Type X strains were engineered to be

drug resistant and crossed with a cat-competent Type II ME49 strain to perform a forward genetic screen to identify the murine virulence loci. Rather than selecting progeny that were double-drug resistant, progeny from these crosses were individually isolated prior to drug selection and then tested to establish whether they were recombinant. Progeny isolated in this unbiased manner allowed for a more accurate determination of the degree to which clones derived from meiosis undergo outcrossing versus self-mating. In four independent genetic crosses, the relative recombination rate was low (approximately 1-3%), rather than the expected 25% rate indicating that in all instances, self-mating was favored. Taken together, the evidence from the Type II and Type X population genetic analyses, as well as from several independent sexual crosses between these two clonal lineages, established that sexual outcrossing is more prevalent than previously envisaged, but that the majority of clones derived from sexual replication closely resemble a single parental type that was amplified by self-mating. And in geographies with predominantly clonal population genetic structures, it is clear that unisexual mating between closely-related strains will be underestimated and not considered as a major contributor to maintaining the clonal population genetic structure. As a consequence, whole genome typing is required to resolve differences between strains within a clonal group, which are preferentially admixing with each other.

Academic advisor, thesis mentor, and primary reader: Dr. Michael E. Grigg

Secondary reader: Dr. Jocelyne DiRuggiero

Thesis committee members (alphabetical): Drs. Orna Cohen-Fix, Jocelyne DiRuggiero, Michael E. Grigg, and Xinzhuan Su

Acknowledgements

This thesis is the culmination of the vast array of support that I received from all of the best people in my life. I cannot thank you all enough for being there for me through this epic journey. This sexy thesis is for you.

I am grateful to Mike Grigg who took me into his lab and taught me how to think and speak like a scientist. Furthermore, without the various members of the Grigg lab I don't think I would have gotten this far. Thanks go especially to Beth Gregg for helping me survive grad school, Patricia Sikorski for telling me when I don't make sense, Cristina Carvalheiro for always being in lab late with me and telling me to go home, Amy Sweeny for keeping science fun, Katie Haman for keeping me grounded, and Dionne Coker-Robinson for being an incredible listener. Additionally, Asis Khan was immensely helpful in explaining how bioinformatics techniques are currently being applied to protozoans.

My collaborators were also invaluable towards accomplishing my work. Thanks to J.P. Dubey (USDA) for his assistance with cat crossing and for always telling me interesting stories. Thanks to Mariam Quinones, Kurt Wollenberg, Kui Shen, and Conrad Shyu from NIAID BCBB for deciphering our problematic haploid genomes with me. Thanks also to Javi Zhang who developed PopNet and helped develop figures from my WGS.

Special thanks to my committee members: Jocelyne DiRuggiero, Orna Cohen-Fix, and Xinzhuan Su who have stuck with me from the beginning. I could not have gotten this far without your guidance and assistance in improving my work.

Similar thanks go to the NIH-JHU GPP without whom I would literally not be writing this thesis. Not only are Orna Cohen-Fix and Michael Lichten great mentors but they were also invaluable to me in becoming an analytical graduate student. Joan Miller (JHU) was invaluable

in keeping all my graduate career officially in order. Likewise, Phil Wang and Phil Ryan (NIH) were completely irreplaceable in their support. Not only did they fix any logistical problem that I came across, but they were invaluable in allowing me the opportunity to pursue my interests by allowing me to teach undergraduates at the NIH and giving me career advice along the way.

My last and deepest thanks go to my family and basically-family. The wonderful Caitlin Dennis, Keri Selzler, and Vivian Nguyen Grimshaw have been like sisters to me for over fifteen years and are always there when I need emotional support. Likewise, David Joy is always willing to explain things to me and Travis Paxton is always there to support me. Rachel Richter and Sam Diener have made me part of their family (and extended families) and are always there to support me no matter what else is going on in their crazy lives.

To my wonderful, beloved husband, who found me before I started this crazy journey and kept me mostly sane through all the late nights in lab hatching oocysts and midnight program errors, who keeps me fed and makes sure I actually sleep sometimes, and who always makes sure I have all the tea I need to keep working. Thank you for always being my worry-wart and partner in crime. I am so glad I sat on your coat.

Thanks also to my cat, Owl, who always knows when I need kitty snuggles and has helped me though the long nights and working weekends.

Finally, thanks go to my parents: To my dad, Allen Kennard, who is always there with a new odd question to answer, a new book to read, and who is always proud of me. To my mom, Laurie Seide, who encouraged me to do what I love even if it's hard, and who is always there for me no matter what I need.

Table of Contents

List of Tables

List of Figures

List of Supplemental Tables

List of Supplemental Figures

Chapter 1 - Introduction

General Overview

An integral paradigm in the evolution of protozoan parasites is the role sexual replication plays in their expansion and adaptation to the environment. Until recently, the extent to which, or even the existence of a sexual cycle and whether genetic hybridization plays a significant role in the life cycle and transmission of protozoan parasites has been under debate (Tibayrenc and Ayala 2014, 2002; Ramirez and Llewellyn 2014) The population genetic structure of many protozoan parasites has been described as clonal, signifying that the limited number of genetic markers used to study their population genetics were in linkage disequilibrium. This assumed clonality has led to the assumption that in protozoa sexual recombination is either extremely rare or impossible, despite that fact that many genes known to facilitate sexual recombination are expressed in these parasites (Cornelissen and Overdulve 1985; Sibley 2009; Pfefferkorn 1981; Pfefferkorn, Pfefferkorn, and Colby 1977). One possible explanation proposed to reconcile protozoan sexual cycles with a clonal population structure, is that inbreeding between more closely related strains within the population has masked the extent to which genetic hybridization is occurring (Heitman 2010). Hence, sexual recombination is missed either because the characterizing genetic marker density is too low or because extensive inbreeding has purified the genomes of the admixture blocks derived from different ancestries. With the advent of whole genome sequencing, it is now clear that genetic hybridization among highly inbred lines can increase structural variation, in the form of chromosomal copy number somy, gene dosage, and localized CNV (copy number variation) above what would be predicted if the parasite line expanded solely asexually. These genetic admixture fingerprints can be used to identify the

extent to which strains within a clonal population are propagated asexually or by sexual inbreeding. Mitotic drift occurring over time also allows previously designated clonal strains to be differentiated by genome sequencing based on low frequency polymorphisms that could be missed using only low-resolution markers. These mitotically derived single nucleotide polymorphisms (SNPs) can be utilized to classify genomic recombination within clades. Determining the contribution of asexual versus sexual propagation is important because sexual recombination of genomes allows parasites to adapt to ever-changing environments and to expand their host range. Although genetic hybridization is both costly to maintain as well as disadvantageous to well-adapted pairs of parasites and hosts, it is nevertheless an advantageous trait to populations, and is ubiquitous across many domains of life.

While bacteria and viruses do not undergo explicit meiotic recombination within their genomes, reassortment of viral genomic segments or horizontal gene transfer within bacteria effectively shuffle the genomes to create new combinations of virulence and survival related genes, which serve a similar purpose to that of meiotic recombination seen among protozoan parasites (Schneider et al. 2011; Bellanger et al. 2014; Vijaykrishna et al. 2011). In fungi, sexual recombination between clones, even within uniparental mating, can recombine genomes, which can lead to antigenic variation as well as improved adaptation when fungi transition into the rapidly adaptable, but slow growing, diploid state from the faster growing haploid state (Heitman 2010; Feretzaki and Heitman 2013). Although sexual cycles for many parasitic protozoa have only recently been discovered, and it is not clear how frequently protozoan sex is occurring in nature, the ability of recombination to impact both the fitness and biological potential of these parasites has been demonstrated. Genetic hybridization in both experimental crosses and natural isolates of *Leishmania* parasites has demonstrated the potential for genetic admixture to generate

progeny that acquire genetic traits from both parents (Rogers et al. 2014; Romano et al. 2014). Genetic shuffling of metabolic pathways has also allowed *Plasmodium* to overcome selection by drug-treatment regimens in endemic regions (Miotto et al. 2013; Claessens et al. 2014). *Toxoplasma gondii* is a highly prevalent and successful protozoan parasite that is considered a generalist pathogen due to its ability to infect, and be propagated by, all warm-blooded vertebrates on the planet. Part of its success is attributed to its capacity to expand either asexually among its intermediate hosts, or to utilize a highly fecund, sexual cycle in its definitive felid host. As such, it is an ideal model organism to investigate the contributing role which sexual recombination plays in the expansion and acquisition of biological traits that impact the fitness of this parasitic pathogen.

Sexual recombination has previously been identified as an advantageous trait within *Toxoplasma,* allowing the parasite to produce progeny capable of escaping host immune defense mechanisms that both parents are unable to withstand. For this haploid parasite, the most parsimonious explanation is that the progeny derived from avirulent parents have recombined alleles from both parental genomes into novel combinations that facilitate resistance to host immunity, allowing certain progeny to exhibit gain of virulence, or biological fitness phenotypes (Grigg, Bonnefoy, et al. 2001). Indeed, subsequent forward genetic experiments mapped the principally responsible genes for this gain of virulence to particular allelic combinations of two parasite protein kinases referred to as ROP5 and ROP18 (Saeij et al. 2006; Taylor et al. 2006; Behnke et al. 2012). In these studies, from the analysis of marker typed genomes of recombinant progeny from sexual crosses between Type III and either a Type I or II strains, ROP5 and ROP18 were predicted to cause the differential virulence between the parental strains and when deleted from the virulent Type I parent abrogated virulence. Later studies found that this

virulence was dictated by the allelic combination of ROP5 and ROP18 inherent in the genome, as only the correct combination (found in Type I strains) allowed ROP5 and ROP18 to bind and phosphorylate host immune-related GTPases (IRGs) to hide the parasitophorous vacuole of *Toxoplasma* from host cellular immunity (Behnke et al. 2011; Fleckenstein et al. 2012). Furthermore, it has been theorized that the expansion and evolution of the mouse IRGs, one of the innate cellular immunity protein classes that defend against *in vivo* parasite growth, is directly correlated to the adaptation of *Toxoplasma* parasites to the murine host (Lilue et al. 2013). Additionally, it has been hypothesized that the differential combinations of toll-like receptors 11/12 (TLR11/12), the primary innate receptors for murine detection of *Toxoplasma*, found across different host species are linked to the expansion of this parasite into new host ranges (Gazzinelli et al. 2014).

However, despite the advantages of genetic recombination, and the existence of genetic mosaics within the *Toxoplasma* population genetic structure (Lorenzi et al. 2016), *Toxoplasma* is generally considered to possess a predominantly clonal population genetic structure (Khan, Dubey, et al. 2011; Howe and Sibley 1995). A limited number of apparently clonal strains dominate the population of *Toxoplasma* globally (Lorenzi et al. 2016; Shwab et al. 2014). This is especially true for the populations circulating in North America and Europe (Khan et al. 2006; Khan, Dubey, et al. 2011; Fux et al. 2007; Khan, Taylor, et al. 2009; Minot et al. 2012). It is highly unusual for a sexually active organism to display such limited genetic diversity of strains as sexual recombination should broadly increase the strain population diversity. One explanation for limited diversity is the retention of certain chromosomes among *Toxoplasma* clones that facilitate an increased ability to expand asexually via oral infection between susceptible intermediate hosts (Khan et al. 2006; Khan, Miller, et al. 2011). Alternatively, evidence from

waterborne outbreaks has established that uniparental mating occurs in nature (Wendte, Miller, Lambourn, et al. 2010), thus a particularly successful clone could have expanded sexually by self-mating and produced the current clonal population structure. Differentiating between these two scenarios is a primary objective of this thesis project.

To explain the widespread abundance of identical genotypes within many species of parasitic protozoa, Tibayrenc and Ayala proposed the clonal theory to explain the simple population genetic structures identified for many protozoan parasites. Simply, the clonal theory argues that recombination within the parasitic protozoa is infrequent, and that the vast majority of isolates within a population are expanding asexually. The molecular determinants that are responsible for generating the clonal population are outside of the speculation of the clonal theory; rather it is an observed phenomenon of population propagation. Theories to explain the clonal theory include the use of too few, poorly resolved genetic markers to resolve closely related strains, drug selection sweeps, founder effect on the population due to a limited number of strains initially genetically founding a sampled area, the natural selection of parasites by susceptible versus resistant hosts, or environmental selection whereby oocyst viability within the environment determines the infection of parasites into new hosts. The clonal theory argues that population clonality is principally propagated by asexual replication and expansion within the population. It argues that sexual replication, if it exists, is not sufficient to break the preponderant clonality observed, and it fails to include uniparental mating as a possible mechanism to explain clonality among haploid parasites. The premise of the clonal theory applied to *Toxoplasma* isolates is that the population is expanding predominantly asexually despite the fact that *Toxoplasma* possesses a highly fecund sexual propagative ability.

The clonal population theory is well supported, especially in North America and Europe where most isolates from humans, livestock, and wild animals appear to be from only four clonal lineages, designated Type I, II, III and X (Khan, Dubey, et al. 2011; Shwab et al. 2014; Dubey et al. 2011). However, there are notable exceptions to the clonal theory. Differences between Type II strains collected from more distant locations, such as Australia and Israel, demonstrate that genetic differences can be resolved between strains within the Type II clonal lineage (Verma et al. 2015; Pan et al. 2012; Parameswaran et al. 2010). Whether this is consistent with the clonal theory cannot be readily determined with the currently available genetic markers and will thus require whole genome sequencing to resolve. Likewise, phenotypic differences between Type I strains have been observed across different laboratories and over time, indicating that the Type I lineage, as with Type II, is either not clonal or has acquired independent mutations from genetic drift during asexual replication, that are sufficient to yield distinct phenotypes (Khan, Behnke, et al. 2009).

These apparent exceptions to the clonal theory necessitate a re-examination of the *Toxoplasma* intra-typic (or within lineage type) population genetic structure to ascertain whether genetic recombination has occurred. Because *Toxoplasma* contains a highly fecund sexual cycle but can also be maintained indefinitely among its intermediate hosts by asexual expansion, it is a choice model system to test the applicability of the clonal theory on protozoan pathogens, as well as to test the extent and contribution of sexual recombination to the population structure and dynamics of this globally-expanded eukaryotic parasite. It is the goal of this dissertation to apply whole genome sequencing and designer scripts to distinguish genetic recombination from genetic drift, and to determine the degree to which sexual recombination has impacted the dominant North American clonal lineages: Type II and Type X. While many studies have focused on the

diversity of this parasite worldwide, none have explored a single clonal lineage in any depth. Hence, the work herein will examine the sufficiency of the clonal theory in *Toxoplasma* and the extent to which sexual recombination is shaping the population genetic structure of this ubiquitous protozoan parasite.

Toxoplasma Lifecycle

Toxoplasma gondii is a eukaryotic unicellular parasitic protist within the phylum Apicomplexa, subclass Coccidian of which other known parasites include *Plasmodium*, responsible for malaria, and *Cryptosporidium*, a well-known livestock parasite. All parasites within the Apicomplexa phylum are defined by their apical complex, a cytoskeletal structure found at the anterior end of the parasite that is comprised of a microtubular network organized into apical rings that tether the inner membrane complex and a series of unique organelles. These organelles, known as rhoptries, dense granules, and micronemes, facilitate host cell penetration and the establishment of their intracellular host environment. *Toxoplasma* is an obligate intracellular parasite, capable of infecting any warm-blood animal, and is capable of establishing infection in any nucleated cell. As a generalist parasite, *Toxoplasma* has a number of unique mechanisms to invade the host cell and manipulate the intracellular conditions toward a favorable environment for the parasite.

Most *Toxoplasma* infections come from ingestion of either tissue-encysted bradyzoites (the transmissible state of the asexual cycle) or environmentally stable oocysts (the transmissible state of the sexual cycle). Both infectious forms rapidly differentiate within the host small intestine into the disseminating stage, known as a tachyzoite, which spreads throughout the host to propagate and establish infection.

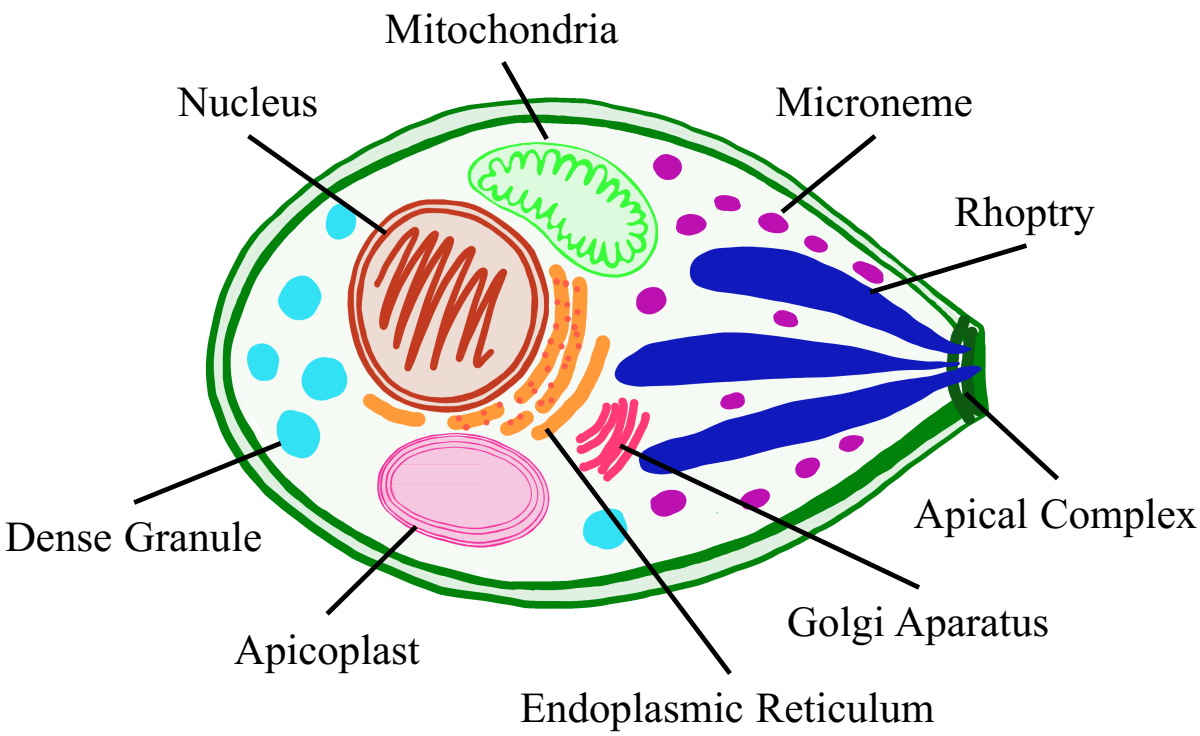


Figure 1-1: Diagram of a *Toxoplasma* tachyzoite with major organelles labeled

Parasite Entry

Toxoplasma is an obligate intracellular pathogen; thus, the parasite must invade a nucleated host cell in order to replicate. While *Toxoplasma* is capable of attaching to and invading any warm-blooded cell that it contacts, not all potential host cells are infected, even after *Toxoplasma* parasites have bound to their surface. “Kiss and spit” is the term coined by Boothroyd and Dubremetz to describe the interactions between the *Toxoplasma* apical complex organelles and a potential host cell prior to invasion (Boothroyd and Dubremetz 2008). During initial contact with the potential host cell, the apical complex secretes specialized protein kinases into the host cell in an ordered sequence of events prior to commitment to invade

(Carruthers and Sibley 1997). If invasion conditions are not determined satisfactory to the parasite, a process that is still undefined, invasion does not commence following the "kiss and spit" injection. These uninfected cells, can be identified via intracellular staining of several secreted *Toxoplasma* proteins. For example, early invasion proteins such as GRA15 are often identified in cells that parasites attach to, but do not enter. These unsuitable host cells are referred to as injected, but uninfected. However, if the *Toxoplasma*-secreted proteins prepare the host cell sufficiently, the parasite begins the host cellular invasion process.

Once invasion of the host cell begins, the *Toxoplasma* parasite forms a tight junction with the host plasma membrane, a structure that has been dubbed the moving junction (Sharma and Chitnis 2013). It is at this point that the proteins from the rhoptry necks and micronemes are secreted, injecting the host with factors such as RON2 and AMA1, which form part of the moving junction that is critical for translocation of the parasite into the host cell (Tonkin et al. 2011). Microneme release is closely followed by rhoptry release, which injects protein kinases (such as ROP5 and ROP18) into the cell. These rhoptry proteins are trafficked to a number of cellular locations and are generally associated with the formation and protection of the parasitic compartment *Toxoplasma* forms, called the parasitophorous vacuole (PV) (Carruthers and Sibley 1997; Sharma and Chitnis 2013). Some of these ROPs are known to attach to the cytoplasmic side of the parasitophorous vacuole membrane (PVM) as it is forming, but others, known as evacuoles, remain within their own smaller vacuoles in the host cytoplasm until later in parasite maturation (Hakansson, Charron, and Sibley 2001). Finally, the parasite secretes the dense granule organelles into the host. While not all of these proteins have a known function, the majority of those investigated facilitate parasite manipulation of the host cellular environment, such as GRA15 which is known to activate host NF-kB to allow a more

productive infection (Rosowski et al. 2011). Only once all three sets of secretory organelles have released their contents into host cells can a successful invasion event be realized, with *Toxoplasma* parasites contained within the PV.

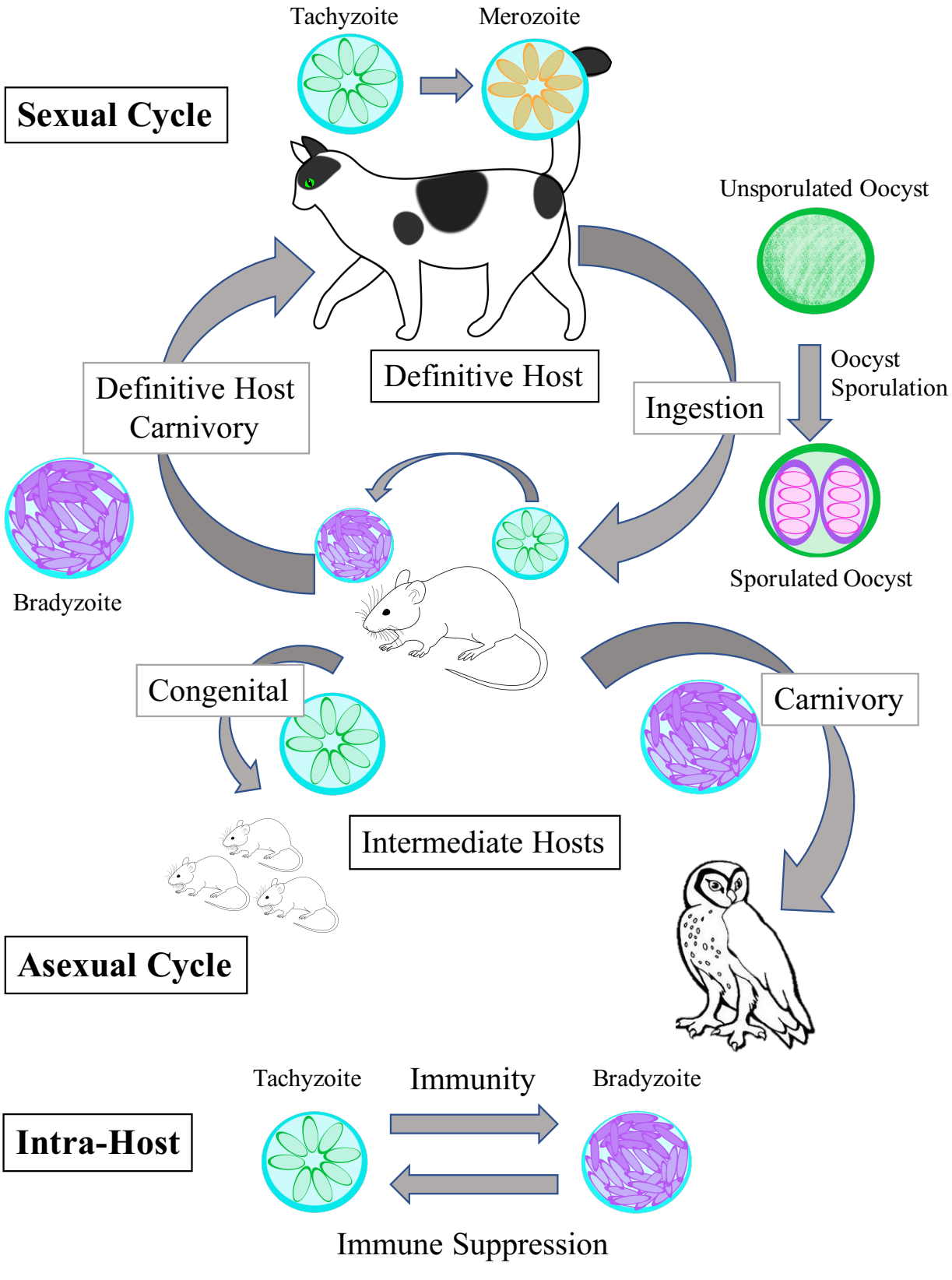
Once the cellular environment is prepared by the secreted proteins, *Toxoplasma* begins active invasion of the host cell. Parasites actively invade the host cell using acto-myosin motors within the *Toxoplasma* to pull the parasite's moving junction and the connected host plasma membrane around the invaginating parasite in a process akin to forced endocytosis. During invasion, the parasite uses the moving junction to exclude host proteins from the stolen plasma membrane in order to ensure parasitic vacuole evasion from detection by innate intracellular host defenses (Mordue et al. 1999).

Following successful invasion of host cells, *Toxoplasma* further modifies its host environment to fulfill its parasitic needs and to avoid host detection. In addition to the evasion of the host using host-self membrane to create the parasite specific vacuole, *Toxoplasma* additionally studs the PVM with its own defensive proteins released from the rhoptry and dense granule organelles during invasion to combat intracellular host defenses. These will be explored in depth later in this chapter. To satisfy its needs as a parasite, *Toxoplasma* recruits key host organelles into close proximity with the PVM in order to more efficiently scavenge host resources. Specifically, the host mitochondria and endoplasmic reticulum are recruited to the surface of the PVM despite *Toxoplasma* having functional copies of both organelles within its own cytosol (Boyle and Radke 2009; Seeber and Steinfelder 2016; Adomako-Ankomah, English, and Danielson 2016; Pernas et al. 2014; Sedwick 2014). Indeed, *in vitro* scavenging of host proteins and key nutrients has been observed (Caffaro and Boothroyd 2011).

Parasite Replication

Following the establishment of a stable infection in a permissive intermediate host cell, the parasite undergoes mitosis and divides asexually within the PV. During this phase of its life cycle, the *Toxoplasma* tachyzoite (named for its distinctive banana shape), replicates its organelles and begins to divide in a unique process called endodyogeny (Dubey and Frenkel 1972). During endodyogeny, the tachyzoite allocates copies of its least complex organelles, such as the Golgi, to opposite sides of the mother cell. Within the multiple-membrane enclosed, complex organelles, such as the nucleus, apicoplast, and mitochondria, the genetic content is internally replicated; the organelle elongates, and then divides (Striepen et al. 2000; Matsuzaki et al. 2001; Vaishnava and Striepen 2006). These organelles then utilize the centrosomes of the dividing parasite to segregate evenly between the two sides of the mother cell. After the organelles have evenly partitioned, two identical daughter cells develop and eventually divide within the mother cell's plasma membrane in a process that greatly resembles bacterial binary fission (Ferguson et al. 2005; Ferguson and Dubremetz 2014). As the process completes, the mother cell's inner membrane complex dissolves and releases the two complete daughter cells into the PV. Although it unknown how the various tachyzoites within the PV coordinate this event, mitotic division within the PV is generally synchronized such that all daughter cells are released at approximately the same time. This mitotic synchronization is often utilized by researchers to determine how many mitotic events have occurred since the host cell was invaded (Ferguson and Dubremetz 2014).

Figure 1-2: *Toxoplasma* lifecycle and replication among hosts
Toxoplasma can infect and replicate in any warm-blooded animal across the globe. It has two major replicative cycles, the sexual cycle which can only occur in the definitive feline host and the asexual cycle which occurs in all hosts infected with this generalist parasite. When a feline ingests an infected intermediate host, bradyzoites from the intermediate host traffic through the cat's digestive tract until they reach the intestine. In the intestine, these bradyzoites invade epithelial cells and differentiate into tachyzoites which spread through the cat. Some of these tachyzoites traffic to the gut enterocytes where they differentiate into merozoites which then undergo gametogenesis to produce male and female gametes. These gametes fuse to produce the immature, unsporulated oocyst which is then excreted in the feces of the feline. In the environment, these immature oocysts sporulate and mature into the highly infectious oocyst. Ingestion of the oocyst by any intermediate host (here a mouse), either via contaminated water or meat, causes these oocysts to excyst their sporozoite progeny in the host digestive tract. These sporozoites then differentiate into tachyzoites in the intermediate host. Carnivory and congenital transmission between intermediate hosts allows *Toxoplasma* to propagate asexually between hosts. Immune resistance within the host causes tachyzoites to differentiate into slow-growing bradyzoites which establish a chronic *Toxoplasma* infection of the host which may be passed to other hosts via carnivory of the primarily infected host. Infection of a pregnant host can lead to parasite-asexual host-congenital transmission. A loss of immune control can allow recrudescence of chronic infections where bradyzoite cysts differentiate into tachyzoites which can cause symptoms of toxoplasmosis in previously asymptomatic hosts.


Tachyzoite
Merozoite
Sexual Cycle
Unsporulated Oocyst
Oocyst
Sporulation
Definitive Host
Carnivory
Definitive Host
Ingestion
Sporulated Oocyst
Bradyzoite
Congenital
Carnivory
Intermediate Hosts
Asexual Cycle
Tachyzoite
Immunity
Bradyzoite
Intra-Host
Immune Suppression

Parasite Egress

Eventually, tachyzoite replication causes the PV to become completely filled with mature tachyzoites. The filled PV takes up most of the available space inside of the host cell and pushes all of the host organelles against the host plasma membrane. At this point, the parasite receives an unknown signal and initiates egress from the host cell via host cell lysis, allowing the nascent daughter parasites to invade new cells and expand the host infection. In response to a yet undetermined egress signal, tachyzoites release a protein known as TgPLP1 to create a series of pores in both the PVM and the host membrane to allow the release of tachyzoites from their initial host cell (Eidell, Burke, and Gubbels 2010; Roiko, Svezhova, and Carruthers 2014). Cell death resulting from host cell rupture is the cause of most clinical manifestations of toxoplasmosis (Dubey and Jones 2008). Released tachyzoites then invade resident cells in proximity to the initially infected cell. This expansion technique makes use of *Toxoplasma's* ubiquitous invasion mechanisms to allow the parasite to spread both locally (via invading intestinal cells in close proximity to the initial infection site) or to disseminate throughout the host by invading immune cells, such as macrophages, that are summoned via the innate and effector arms of the immune system to control *Toxoplasma* infection. This process facilitates dissemination of the *Toxoplasma* infection to more distant parts of the infected host, such as the lungs and brain.

Parasite Encystment

To survive and establish long-term, chronic infections within a host, the tachyzoite can undergo a developmental switch from its fast growing, actively replicating form, to the slow growing bradyzoite form, which has the capacity to encyst inside a host cell. Proliferating tachyzoites in an immune-competent host are eventually detected by the innate and effector arms of the immune system. Innate immune cells activated by IFN-γ generate an immune response against tachyzoites that can kill off most of the parasites circulating during acute infection (Suzuki et al. 1988). In certain cell types, such as striated muscle and the central nervous system, *Toxoplasma* escapes this immune activation by differentiating into its chronically infectious bradyzoite form, which derives its name from the Greek word "brady" that means slow. The exact trigger for differentiation is unknown, although it is likely a combination of host immune stress in response to tachyzoite growth and intracellular signaling within the PV (Skariah, McIntyre, and Mordue 2010; Soete et al. 1993; Bohne, Heesemann, and Gross 1994). Once differentiated, bradyzoites replicate within the PV at a much lower rate than their tachyzoite predecessors in order to remain largely undetected by the host immune system (Watts et al. 2015). In the majority of infected hosts, *Toxoplasma* persists because the immune response induced by tachyzoites, which successfully limits tachyzoite proliferation, is unable to clear bradyzoite tissue cysts. However, these bradyzoite cysts require a steady state of host immune control to remain quiescent. Latency of the infectious bradyzoite cysts not only allows *Toxoplasma* to be transmitted to any carnivorous intermediate host to complete its asexual cycle, it also complicates the survival of chronically infected hosts, since any perturbation that causes immune suppression can lead to recrudescence of *Toxoplasma*

infection, as is seen in pregnant women, organ transplant, and AIDS patients (Brabin 1985; Ayi et al. 2016; Derouin, Pelloux, and Parasitology 2008; Dubey and Jones 2008; Gross et al. 1997).

Parasite Sexual Cycle

When an intermediate host with bradyzoite tissue cysts is consumed by the definitive feline host, *Toxoplasma* will enter its sexual cycle. Bradyzoite cysts, which are resistant to stomach acid, pass through the feline stomach and traffic to the small intestine where they infect the feline intestinal epithelium. Although asexual replication within the guts of a variety of intermediate hosts is well characterized, the developmental switch that commits the parasite to differentiate into a merozoite, instead of a tachyzoite, and enter its sexual cycle within the feline lumen is poorly characterized, beyond a basic ultrastructural description (Ferguson, Hutchison, and Siim 1975). As such, characterization of the distinct life cycle signals and stages present in the sexual cycle is largely a black box in the *Toxoplasma* field. Within cat epithelial enterocytes, the vast majority of bradyzoites differentiate into merozoites rather than the default tachyzoite produced in all intermediate hosts (Hehl et al. 2015). Merozoites then undergo asexual expansion, much like their tachyzoite counterparts, for a non-specific number of mitotic replications. At some point, typically 3-5 days post infection, merozoites differentiate into the sexual stages of *Toxoplasma.* Each haploid merozoite differentiates into either haploid microgametes (male gametes) or a single haploid macrogamete (female gamete) within its host cell (Ferguson 2002). The microgametes resemble eukaryotic spermatozoa but are bi-flagellated, whereas macrogametes closely resemble eukaryotic oocytes and similarly

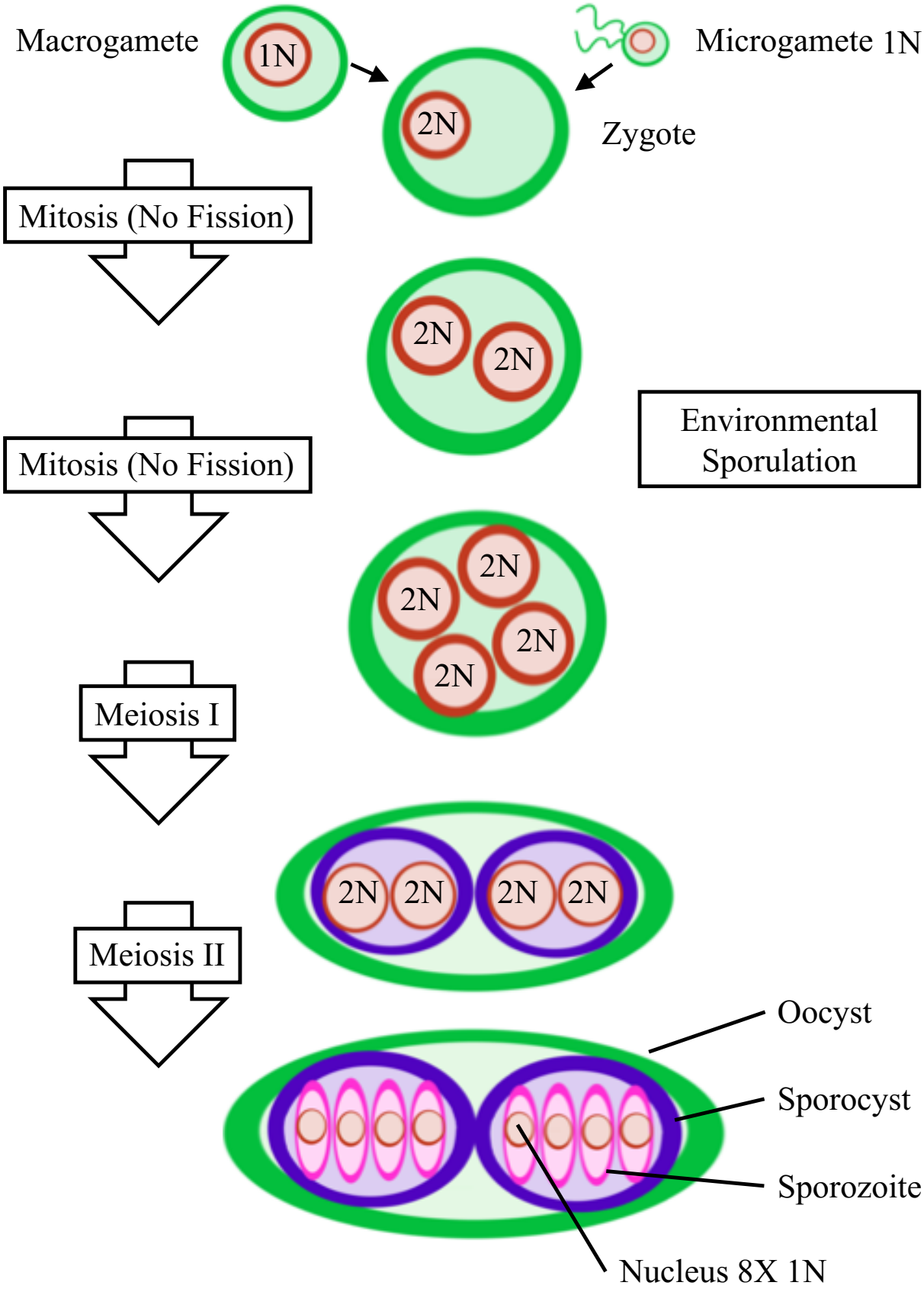
bear maternally inherited organelles within them, such as the mitochondria and the apicoplast organelle that is present in a majority of Apicomplexan parasites (Ferguson 2002). Macrogametes are also known to produce a thick outer wall that later becomes the oocyst outer wall and which makes microscopic interrogation of the sexual stages particularly difficult (Ferguson and Dubremetz 2014).

It has been determined empirically that gametes are produced in a roughly one to one ratio, but there is no current evidence about gamete sexual determination, gamete attraction to one another, where fertilization occurs, or why the sex ratio differs drastically from that of other apicomplexan parasites. It has been suggested previously that the ratio of micro-to-macrogametes produced would support a model whereby *Toxoplasma* sexual recombination is biased toward self-mating (Ferguson, Hutchison, and Siim 1975; Ferguson et al. 1974; Read et al. 1995). Whether fertilization occurs intracellularly, or extracellularly is a matter of debate. Regardless, gametes are found within the lumen of the feline gut where it is postulated that microgametes receive a chemical signal to locate the unfertilized macrogamete. The gametes then fuse to create the diploid oocyst that can be formed from gametes from a single strain infection to create a clonal oocyst, or if multiple strains of *Toxoplasma* have infected the feline, then mixed-strain recombinant oocysts may be formed. While there is some evidence that not all strains of *Toxoplasma* can self-mate during sexual expansion, there are not enough sexual crosses to definitively pin-point the cause of this (Fux et al. 2007). The oocysts are then excreted from the cat in its feces (Ferguson 2002).

Once excreted into the environment, sporulation of the oocyst occurs. The diploid oocyst (2N) undergoes mitosis twice without fission to produce a single zygote with four identical 2N genomes (8N total). This DNA replication is followed by meiosis I where sexual

recombination of the genomes occurs to produce two separate sporocysts, each with two 2N recombined genomes (4N total per sporocyst), encapsulated inside of the oocyst outer wall. The 8N oocyst containing two 4N sporocysts then undergoes division in meiosis II to create four independent 1N sporozoites inside of each sporocyst, to create a total of eight 1N sporozoites inside two sporocysts inside of the single oocyst. In total, eight haploid sporozoites, all independently recombined and assorted from one another, are sporulated and matured inside a single oocyst after excretion from the definitive feline host. Once the oocysts are sporulated they are capable of infecting any warm-blooded vertebrate. Oocysts are incredibly environmentally stable to both physical and chemical assault and have been shown to be viable in various conditions for years without losing infectivity (Yan et al. 2016; Lindsay and Dubey 2009; Jones and Dubey 2010; Lindsay, Blagburn, and Dubey 2002; Dubey 1998). The oocysts are transmitted either through contaminated water or food and pass through the host gut until stomach acid and lowered carbon dioxide levels begin the process of excystation and the sporozoites are released into the host small intestine (Freyre and Falcon 2004; Tilley et al. 1997). At this point, sporozoites rapidly differentiate into tachyzoites and begin the asexual cycle of replication within the intermediate host.

Figure 1-3: Oocyst sporulation after environmental excretion
Oocyst zygotes are formed by micro and macrogamete fusion in the feline definitive host. These immature oocysts are then excreted into the environment where they sporulate. The diploid oocyst (2N) undergoes mitosis twice without fission to produce a single zygote with four identical 2N genomes. These four 2N genomes undergo meiosis I to form two sporocysts, each with two 2N recombined genomes. Each of these sporocysts genomes undergoes meiosis II to form four independent 1N sporozoites inside of each sporocyst inside of the oocyst. In total, eight independently recombined and assorted, haploid sporozoites are contained in a single oocyst after excretion from the definitive feline host.


Macrogamete
1N
Microgamete 1N
2N
Zygote
Mitosis (No Fission)
2N
2N
Environmental Sporulation
Mitosis (No Fission)
2N
2N
2N
2N
Meiosis I
2N
2N
2N
2N
Meiosis II
Oocyst
Sporocyst
Sporozoite
Nucleus 8X 1N

Parasite Life Cycle Transmission

Toxoplasma can be transmitted to a new host via its asexual life cycle when tachyzoites are passed vertically to a developing fetus following infection of a non-immune pregnant female, or from immune females that fail to maintain sufficient immunity to control the recrudescence of latent bradyzoite cysts. The process of vertical transmission (via congenital infection) is one way in which a single clone can be expanded asexually to maintain the parasite life cycle in the absence of the definitive felid host (Hide 2016). Alternatively, because bradyzoite cysts are orally infectious, it is also possible to horizontally expand a single strain asexually across a broad intermediate host range via carnivory and scavenging (Fux et al. 2007).

Parasite sexual transmission occurs when tissues harboring bradyzoite cysts from infected intermediate hosts are consumed by the definitive felid host. The fecundity of the sexual cycle is profound. A single feline is capable of expanding the parasite 10^7-fold once infected. Oocysts defecated from the definitive felid host are highly infectious. If a cat is infected by a single *Toxoplasma* genotype, the parasite can undergo its full sexual cycle by self-mating, as a single strain is capable of producing both micro and macrogametocytes and no obvious mating-type locus is present in the *Toxoplasma* genome. This form of sexual transmission is very difficult to distinguish from parasites that are expanding asexually, as all progeny from the sexual cycle of a feline infected with a single strain of *Toxoplasma* are virtually identical to the parental strain. However, if two or more genetically distinct strains infect a single feline, the same exponential biological expansion occurs, but now many of the progeny from this outcross are recombinant genetic admixtures between the two parental

genomes. Outcrossing has been shown previously to greatly impact the biological potential of the parasite because it brings alleles from both parental strains into new combinations that impact virulence or the capacity to inactivate host immune signaling networks, as has been described previously (Grigg, Bonnefoy, et al. 2001; Saeij et al. 2006; Taylor et al. 2006). Furthermore, because oocysts are generally highly infectious and are readily dispersed (in water supplies, for example), they have a greater potential to infect a wider range of hosts (including herbivores and carnivores) than the 1:1 asexual transmission rate from carnivorous consumption of bradyzoite infected prey (Dubey 2001).

Due to its capacity to utilize both its sexual and asexual cycles, *Toxoplasma* has a vast array of replication techniques that help this parasite dominate as one of the most prevalent parasites in both humans and animals on the planet. Sexual self-mating has been previously observed for other parasitic pathogens, including fungi and kinetoplastids (Heitman 2010; Sun et al. 2014; Iantorno et al. 2017). Unlike other heteroxenous coccidians, *Toxoplasma* is distinct among the apicomplexan parasites in having the replicative flexibility to sexually recombine without the necessity to do so to propagate. Furthermore, *Toxoplasma* is not limited to either outcrossing or self-mating when it sexually recombines, which could lead to the underestimation of the occurrence of sexual replication within the *Toxoplasma* population structure because it is very difficult to differentiate transmission of clones from the asexual cycle (bradyzoite cysts) from that of oocysts derived from sexual self-mating. To understand which of these two cycles is responsible for the widespread dissemination of parasite clones, it will be necessary to develop tools that can distinguish the extent to which infection by bradyzoites or sporozoites occurs in nature, or to determine whether sexual self-mating generates sufficient structural variation -in the form of copy number variation (CNV),

alterations in ploidy, or other hallmarks of meiosis- that can be used to distinguish between these two mechanisms of transmission.

Prevalence of *Toxoplasma* in Humans

Humans are the most widely surveyed of all intermediate hosts of *Toxoplasma*. Prevalence rates vary across different geographies, and infected people are found on every continent on the planet, including Antarctica (Dubey 2010). While human infection represents the majority of infection information obtained for a single host, especially across both geographical and longitudinal studies, humans are ultimately an unproductive host for *Toxoplasma.* Humans are neither generally prey for large felids, which would facilitate propagation of parasite clones via the sexual cycle, nor are they considered prey for other carnivores and scavengers to facilitate asexual transmission. Furthermore, humans do not readily transmit *Toxoplasma* vertically via congenital infections (as discussed later). This evidence suggests that humans are a dead-end host for parasite transmission (Li et al. 2014). Infection rates by *Toxoplasma* vary widely across the parasite's range based on geographic and environmental compatibility to oocyst infection (for example, frozen climates are less likely to have oocyst-initiated infections), cultural proclivities that can increase contaminated food transmission (such as the consummation of raw meat or unsterilized water), and human immune susceptibilities.

Across the human population, *Toxoplasma* endemic infection rates vary widely from 4% in Korea to 92% in the Amazon regions of Brazil (Ryu et al. 1996; Figueiró-Filho et al. 2005). These infections can stem from a variety of sources, including meat containing

bradyzoite cysts and unwashed vegetables contaminated with soil oocysts (Tenter, Heckeroth, and Weiss 2000; Boyer et al. 2011; Mead et al. 1999). However, perhaps the most prevalent and least recognized source of human *Toxoplasma* infections comes from the contamination of water sources with cat-derived oocysts (Boyer et al. 2011; Demar et al. 2007; Jones and Dubey 2010; Torrey and Yolken 2013; Vaudaux et al. 2010). This contamination is greatly facilitated by the high environmental stability of oocysts. Oocysts have been found to be environmentally stable across a wide range of environmental conditions, however transmission to human hosts is much higher under certain conditions (Dubey 1998). In a large-scale analysis of human clinical data, the McLeod group determined that the infection of humans by *Toxoplasma* was significantly increased in warm, damp areas over that of drier mountainous areas (Lykins et al. 2016). This transmission variation may partially explain why endemic *Toxoplasma* infections are much higher in warm, humid areas such as the North American south and the Amazon regions of Brazil over that of cooler and drier northern North America and Europe where conditions are less favorable to oocyst environmental stability. However, this report also points out that seroprevalence of *Toxoplasma* may be underestimated due to a combination of asymptomatic infections and under/miss-diagnosis of *Toxoplasma* infections as other common diseases. These problems are especially prevalent in uninsured human populations, which are both more likely to live and work in endemic regions, which increases exposure to the parasite, and who are less likely to seek treatment unless symptoms become severe (Yan et al. 2016).

Humans may in fact be more resistant hosts to *Toxoplasma* than other intermediate hosts since they rarely succumb to the parasite without extant immune suppression. Most infections in humans are asymptomatic and thus go unreported and untreated. Indeed, the detection of *Toxoplasma* in immune suppressed communities is invariably higher than it is in

the general population, and quite possibly represents the true prevalence of *Toxoplasma* infection in humans (Ayi et al. 2016; Herrmann et al. 2014; Lykins et al. 2016). Moreover, even symptomatic infections can go unrecognized as the symptoms of *Toxoplasma* resemble those of more common diseases such as the common cold and influenza (Lykins et al. 2016).

Strain differences in *Toxoplasma* account for some of the diversity seen in human parasitic infections (Rico-Torres, Vargas-Villavicencio, and Correa 2016; Xiao and Yolken 2015). However, the differences between strain infections are difficult to parse because most human infection data is systematically limited by a lack of human host genetic characterization and diversity, and the limited ability to adequately sample *in vivo* parasites to characterize genetics. Much of our knowledge of human infections comes from one of three limited subsets: symptomatic immune-competent patients, symptomatic immune-compromised patients (such as those with AIDS or organ transplants), and pregnant women (usually asymptomatic) surveyed prenatally as a state-required, transmission preventative measure. Not only do these survey methods miss most asymptomatic infections, but most offer only rudimentary characterization of the strain the patients are infected with. The majority of infections in humans are characterized by either serological reactivity or via PCR amplification of parasite DNA from the host blood (Zhang et al. 2016; Wyrosdick and Schaefer 2015). Both methods come with severe limitations beyond the limited sample of humans surveyed. For serological testing, *Toxoplasma* parasites are strain typed based on the reactivity of antibodies in the serum to peptides designed to segregate only the three most common clonal types of *Toxoplasma* (Vaudaux et al. 2010; Shobab et al. 2013; Kong et al. 2003; McLeod et al. 2012). While this reactivity is seen in both acute and chronic infections it can be lacking in immune compromised infections due to immune inability to produce a response to the parasite peptides

(Schoondermarkvandeven et al. 1993). PCR amplification of the parasite DNA from the blood of human infections is problematic as this detection requires free parasites in the blood stream. As most *Toxoplasma* parasites are tissue resident rather than blood-born, these free parasites are often in limited supply, and only multi-copy genes can be amplified, which are genetically less informative than the strain specific diagnosis that is only possible using highly polymorphic single-copy genes (Piergili Fioretti 2004).

It has been shown that human infections in areas of endemic Type II *Toxoplasma* regions trend mainly toward asymptomatic infections, which are mainly detected during the mandatory prenatal screenings of pregnant women as well as from recrudescence of *Toxoplasma* in immune compromised humans (Ayi et al. 2016; Herrmann et al. 2014). In comparison, human infections in areas such as Brazil, that possess a more diverse population genetic structure of strains, tend to display more severe symptomatic infections that range from protozoan encephalitis to ocular toxoplasmosis (Demar et al. 2007; Ferreira et al. 2008; Grigg, Ganatra, et al. 2001; Jones and Dubey 2010). Indeed, ocular *Toxoplasma* infection has been closely associated with more diverse strains which are less well-controlled by the human immune system and are more capable of breaching the blood-brain barrier (Fekkar et al. 2011; Grigg, Ganatra, et al. 2001; Shobab et al. 2013; Grigg, Dubey, and Nussenblatt 2015). Information on the origin of the parasite infection (as opposed to parasite collection) as well as human genetics, are often not documented, making information derived from human infections tenuously characterized at best. These limitations create difficulties in separating the strain polymorphism differences linked to human disease from the differences in human immune system susceptibility that these strains encounter. However, this interaction between

Toxoplasma parasite strains and the host immune system is more well-understood in other, better studied intermediate hosts (*i.e.,* rodents) of *Toxoplasma,* discussed later in this chapter.

Toxoplasma infection is mainly problematic when it is not tightly controlled by the host's immune system, such as is the case in the immune compromised. During organ transplantation, it is now common to treat seronegative recipients prophylactically to limit a potential infection from a seropositive donor (especially if donor testing was not done) (Paya et al. 2012; Derouin, Pelloux, and Parasitology 2008). *Toxoplasma* is also an immense problem in HIV endemic regions not only due to problems associated with primary infection in immune-compromised hosts, but also for those *Toxoplasma*-exposed HIV infected individuals who develop AIDS. This patient group is particularly susceptible to a recrudescent infection, in which the latent bradyzoite stage reactivates to produce an acute symptomatic infection in the absence of protective immunity (Gross et al. 1997). Recrudescence is especially challenging in areas with high levels of immune-compromised humans as there is no intervention that yields a sterile cure of the *Toxoplasma* infection. While acute infections can be treated, the bradyzoite state is refractory to all known drug combinations and requires an intact immune system to control its ability to recrudesce and cause toxoplasmosis.

Immune-competent seronegative pregnant women are at high risk of transmitting *Toxoplasma* to their developing fetus. Congenital transmission of an acute infection may occur without the mother displaying characteristic symptomology (Kim 2006). Congenital transmission is more frequent in areas with low endemic levels of *Toxoplasma,* because there are more at-risk seronegative women. In these areas treatment focuses on detecting and treating primary infections in women (Rico-Torres, Vargas-Villavicencio, and Correa 2016). In highly endemic areas, where a significant proportion of women of child-bearing age have been

previously exposed to the parasite, the risk of congenital transmission is lower due to existing maternal immunity. However, in regions where *Toxoplasma* genetic diversity is high, there is a risk of a seropositive mother contracting a secondary infection that breaks primary immunity and allows congenital transmission of the secondary, genetically different *Toxoplasma* strain to the fetus (Elbez-Rubinstein et al. 2009). Surprisingly, while adult human symptoms correlate to the infecting strain's genetics, fetal disease severity is correlated primarily with the trimester of infection, with an inverse correlation between the likelihood of infection and the potential hazard to the fetus (late-term infections being correlated with less symptomatic congenital toxoplasmosis and earlier infections correlating more frequently with fetal abortion) (Boyer et al. 2011; Rico-Torres, Vargas-Villavicencio, and Correa 2016).

Congenital *Toxoplasma* infection is often undiagnosed due to the lack of maternally inherited or fetus-produced antibodies against *Toxoplasma* for those fetuses infected later in the third trimester. Many congenitally infected infants only display symptoms later in life, often around puberty, with ocular toxoplasmosis as the most common presentation (Rico-Torres, Vargas-Villavicencio, and Correa 2016). However, ocular toxoplasmosis is not solely derived from congenital infections, and can occur during, or shortly after, acute disease. Generally, ocular toxoplasmosis is self-limiting and most cases are not treated unless the resultant lesions begin to cause progressive blindness (Harrell and Carvounis 2014). Unfortunately, treatment options are limited and often require direct ocular delivery of chemo-prophylactic agents to limit parasite growth. Even with treatment, ocular lesions remain, and clinical trials have not shown a strong correlation between treatment and prevention of vision loss, although treatment may limit future recrudescence that can trigger additional ocular scarring.

Prophylaxis is more effective in preventing transmission of *Toxoplasma* into seronegative or immune-compromised humans, but treatment can only suppress acute infections and does nothing to clear chronic infections or tissue cysts. Furthermore, the pyrimethamine and sulfonamide drug combinations available to treat *Toxoplasma* infection have a number of undesirable side effects, in addition to not providing a sterile cure for the infection (Soheilian et al. 2005). Ideally, all possible at-risk humans would be vaccinated against *Toxoplasma* infection, but there are currently no human-approved *Toxoplasma* vaccines (Liu, Singla, and Zhou 2012). Instead, *Toxoplasma*-endemic regions focus on prevention of primary infections. This effort has included the addition of mandatory prenatal screening in areas such as France, Germany, and Brazil (Kim 2006). These efforts appear to have reduced the transmission of parasite congenital infections (Avelino et al. 2014).

However, the majority of *Toxoplasma* infections found in humans have zoonotic origins (Schluter et al. 2014). While efforts can be made to keep food safe (via safe handling practices of meats and vegetables) and water may be stringently filtered to exclude oocysts, the zoonotic transmission of *Toxoplasma* to humans is unlikely to be contained any time soon. Due to the inherent lack of information from human hosts who are often asymptomatic, and are infrequently tissue sampled, perhaps the best way to appreciate the impact of *Toxoplasma* on the human population is through the study of the intermediate and definitive hosts that serve as zoonotic reservoirs for humans. This is especially true of oocyst-derived infections where sexual replication in the feline host allows for the genetic admixture and expansion of parasite clones that may have distinct biological potentials, including a greater capacity to infect and cause disease in humans.

Zoonotic Potential of a Generalist Parasite

Toxoplasma is considered a generalist parasite which possesses a heteroxenous lifecycle and is capable of infecting any species of warm-blooded animal on the planet, whereas other apicomplexan parasites are generally more host restricted. For example, although the genus *Sarcocystis* has a broad intermediate host-range, the majority of *Sarcocystis* species are specialized, each possessing a definitive host and a restricted (to one or a few) intermediate host species (Tenter 1995). *Plasmodium* also displays a limited host range. In general, a single *Plasmodium* species can only infect a small subset of mosquito species and an equally small number of intermediate host species (Molina-Cruz and Barillas-Mury 2014). Unlike these other apicomplexan parasites, the single species of *Toxoplasma* can utilize any warm-blooded animal as its intermediate host, and any felid as its definitive host. In addition, *Toxoplasma* has a remarkably flexible transmission cycle wherein it can undergo or bypass its sexual cycle because bradyzoites are orally infectious and are not terminally differentiated or committed to the sexual cycle. This allows *Toxoplasma* to maintain a high infection prevalence in its globally wide host range, where it infects animals across every continent and in every ocean on the planet.

The progression of toxoplasmosis in animal hosts mirrors that seen in humans. More strains have been isolated from animals due to the ease of animal tissue sampling for strain isolation. While the ability to isolate strains is higher in animals, most information from animals comes from post-mortem analysis of predominantly symptomatic infections. Additionally, domestic animals are more closely sampled, especially with regard to longitudinal and known-age of host studies than are wild animals. Wild animals tend to be

sampled opportunistically, for example when caught during hunting, or when they are inherently symptomatic. This biased sampling lead to the hypothesis in the field that wild strains of *Toxoplasma* are more likely to cause disease. However, a large portion of the "wild" animals diagnosed with clinical disease are from captive animals in domestic locations, such as zoos, and there is an increasing body of evidence that wild strains of *Toxoplasma* are just as likely as domestic strains to be avirulent in their respective hosts (Thompson 2013; Thompson, Lymbery, and Smith 2010).

Perhaps the most significant host for *Toxoplasma* transmission is the cat. Every felid species is considered a definitive host and is capable of supporting the full sexual life cycle, including the ability to produce highly stable oocysts by outcrossing or same strain self-mating. This transmissible stage can infect any warm-blooded domestic or sylvatic animal that ingests it. The global *Toxoplasma* seroprevalence in domestic cats is estimated to be between 30-40%, however, as with humans, seroprevalence may underestimate the actual infection rate due to the low-level infection detection limit, especially using older tests (Schluter et al. 2014; Lopes et al. 2014). Furthermore, some labs have hypothesized that seroprevalence in wild cats is closer to 100% by adulthood, reaching these numbers as the animals aged (Lopes et al. 2014). Indeed, transmission rates in some intermediate hosts, such as the Southern sea otter, may be better correlated to wild felid populations rather than the domestic felids closely associated with humans (VanWormer et al. 2016).

Livestock are also an especially prominent, well monitored zoonotic source for human infections. Not all domestic livestock exhibit the same rates of seroprevalence, consistent with differential exposure and infectivity rates across meat animals, as well as different immunological defenses between animals against this generalist parasite. A meta-analysis of

food animals in the United States found the following seroprevalence rates in non-confinement pigs (31.0%), goats (30.7%), non-confinement chickens (24.1%), sheep (22.0%), confinement breeding pigs (16.7%), and confinement market pigs (5.6%) across the United States. The analysis also surveyed both cattle and confined chickens but found that both possess low seropositive rates (Guo et al. 2016). Several common factors influence infection rates of these species, including exposure to the environment and age of the animal surveyed. In all species of livestock assessed, seroprevalence rates increase linearly with animal age at time of survey (although breeders generally have higher seropositive rates than market stock), displaying a time-dependent correlation between exposure to potentially oocyst-contaminated water and feed, and seropositivity rates of all livestock species (Guo et al. 2015).

The introduction of more stringent farming and fattening practices (reducing exposure to infected feed and environments) has significantly reduced the seroprevalence rates in swine (Tenter, Heckeroth, and Weiss 2000). As shown above, wild swine have the highest seroprevalence for *Toxoplasma* followed closely by organic, free-range farmed swine who have access to outdoor ranges. The lowest seroprevalence in swine has been observed in indoor-raised and fattened swine that were taken to market at six months of age, although it is noteworthy that animals, like humans, have an increased likelihood of infection with Toxoplasma as they age. This becomes especially apparent when comparing samples from livestock and wild (or sylvatic) animals. While most domestic animals are surveyed systematically either at a set life stage or after culling for food production, most wildlife are surveyed from dead specimens either from hunting catch sampling or necropsy of animals found dead, which may not represent the most-healthy animals. Thus, while domestic animals are surveyed in a manner biased towards finding early or congenital infections, wild animals

are generally biased toward only the most symptomatic infections, which have resulted in the death of the animals.

Despite the wide variety of intermediate hosts that *Toxoplasma* infects, certain hosts appear to be more susceptible to specific infectious stages (*i.e.*, bradyzoite versus sporozoite). For instance, felines require a much lower dose of bradyzoites to create a productive infection than mice. Inversely, mice require significantly fewer sporulated oocysts to produce infection than the definitive felid host (Dubey 2001). Likewise, strain type susceptibility is another variable that affects host range. For example, birds are more likely to harbor Type I strains than rodents, whereas livestock are more likely to be infected with Type II or III strains than symptomatic wildlife, while wild animals in North America are most likely to be infected with Type X (Wendte, Gibson, and Grigg 2011). Host selection based on genotype is not uncommon among the Apicomplexa. As discussed previously, different species of *Plasmodium* and *Sarcocystis* are restricted to specific host species to complete their life cycles, and different species of *Eimeria* are known to partition *in situ* to different regions of the gastrointestinal tract within the same intermediate host (Mackinnon and Read 2004; Tenter 1995; Walker et al. 2013).

Previous work has suggested that two distinct pools of *Toxoplasma*-infected hosts exist that comprise distinct domestic and sylvatic transmission cycles with each host pool infected with largely non-overlapping *Toxoplasma* genotypes (Wendte, Gibson, and Grigg 2011). However, when host overlap occurs between the two transmission cycles, strains usually restricted to one host pool have been known to cause outbreaks, such as Type II strains in marsupials and sea otters (Conrad et al. 2005; Parameswaran et al. 2010). It is thought that these cross-cycle infection strains are more symptomatic due to a lack of adaptation between

the host immune system and the infecting *Toxoplasma* genotypes (Wendte, Gibson, and Grigg 2011). In Brazil, where both domestic and sylvatic cycles have been more thoroughly surveyed, sylvatic cycle parasites are typically more genetically diverse than those found infecting domestic livestock. However, based on isolated parasites, the amount of cross-cycle transmission varies by location and hosts surveyed (Dubey et al. 2006; Silva et al. 2014; Vitaliano et al. 2014). Overall, it remains unclear how much host and parasite mixing definitively occurs between the domestic and sylvatic cycles as urban and rural hosts are rarely surveyed in overlapping geographical regions. Nowhere is the potential for outbreaks from novel genotypes more apparent than in "adapter" species of wildlife (e.g. raccoons and foxes) that have adapted to live in human environments that encroach upon wild environments, and which seem to can carry infections from diverse strains of *Toxoplasma* found infecting sylvatic animals into human contact (Thompson 2013).

Population Genetics Indicate *Toxoplasma* is a Single Species with Multiple Diverse Clades Worldwide

Initial studies classified *Toxoplasma* as a single species that is dominated by three clonal lineages (referred to as Types I, II, and III). This work was based on the application of a limited set of PCR-RFLP markers, isoenzyme analyses, and parasite murine virulence (Darde, Bouteille, and Pestre-Alexandre 1992; Sibley and Boothroyd 1992). Although their genomes differ by less than 2% between clonal strains, each clonotype can be differentiated by specifically designed RFLP markers that target known polymorphisms between the clonal lineages. Additionally, each clonotype exhibits a largely unique and distinct virulence

phenotype in mice by which they were initially classified, which correlates well with genomic interrogations. Type I strains cause lethal infection in all laboratory mice, even at low inoculums, often killing with a single tachyzoite. Type II and III strains are much less virulent and require a dose of 2000 or $>10^5$ tachyzoites respectively to cause lethal infection in half of the mice infected (Sibley et al. 1992; Pfefferkorn and Pfefferkorn 1979). While all clonotypes can infect any warm-blooded vertebrate, they partition differentially among susceptible intermediate host populations in nature. Specifically, in humans the majority of infections are from Type II strains that cause chronic, asymptomatic infections, whereas Type I infections are rare but are associated with more severe eye disease (Boothroyd and Grigg 2002; Grigg, Ganatra, et al. 2001). Birds in contrast, are most commonly infected by Type I strains found as asymptomatic infections (Miller and Grigg, unpublished). Type III infections are highly prevalent in domestic animals and wildlife but are not typically found in human infections (Grigg and Sundar 2009).

Following the initial studies that defined the three archetypal lineages of *Toxoplasma*, a suite of 11 PCR-RFLP markers were developed and applied to determine the true extent of *Toxoplasma* global genetic diversity (Su, Zhang, and Dubey 2006). Further studies have applied microarray, microsatellite sequencing, EST sequencing, and finally whole genome sequencing of 62 globally distributed *Toxoplasma* isolates to further refine the molecular characterization of the parasite's population genetic structure (Ajzenberg et al. 2010; Blackston et al. 2001; Khan et al. 2005; Su, Zhang, and Dubey 2006; Boyle et al. 2006; Lorenzi et al. 2016; Minot et al. 2012).

Despite *Toxoplasma's* highly fecund sexual cycle and its ability to recombine within any feline host, it has a surprisingly clonal-dominated population structure across many parts of the globe. In both North America and Europe, the population structure is dominated by the Type II

lineage (Shwab et al. 2014; Su et al. 2012; Wendte, Miller, Lambourn, et al. 2010; Howe and Sibley 1995; McLeod et al. 2012). Australia too exhibits a population that resembles a drifted version of the North American and European population structure (Parameswaran et al. 2010; Pan et al. 2012). Additionally, although limited data has been collected from the regions, *Toxoplasma* collected in China and East Asia also indicates the presence of a clonal population structure, albeit one comprised of lineages considered atypical in both North America and Europe (Dubey, Zhu, et al. 2007; Dubey, Huong, et al. 2007; Zhou et al. 2009). In fact, the only geographic region where the population structure of *Toxoplasma* is as expected for a parasite capable of sexual recombination is South America. In South America, the *Toxoplasma* population displays high genetic diversity and strain admixture, with no one genotype dominating. This panmictic population genetic structure exhibits a heterogeneous array of genetically inter-related strains that would be expected from continuous interbreeding between closely-related but genetically distinct lineages of *Toxoplasma* (Rajendran, Su, and Dubey 2012; Shwab et al. 2014).

Explaining the origin and transmission dynamics for the panmictic population structure in South America compared to the clonal population structure in North America and Europe has been a matter of debate and has generated many opposing theories to explain how two apparently opposite population genetic structures co-exist in nature. While other parasites have shown evidence of clonal lineage sweeps, this is usually due to a selective advantage such as drug selection across the region surveyed, as with chloroquine selective sweeps and artemisin-resistance in *Plasmodium* (Ariey et al. 2014; Miotto et al. 2013; Wootton et al. 2002). However, no drug sweep could explain the clonal dominance of just a few strains in North America and Europe as anti-parasitic drugs for *Toxoplasma* do not provide a sterile cure of the

parasite and thus would not allow an efficient genetic sweep. Nevertheless, biased sampling may have skewed the sampled population genetic diversity toward clonality, as most samples collected in Europe and North America are derived from a limited host subset of domestic animals and symptomatic humans (Wendte, Gibson, and Grigg 2011). One hypothesis to explain the clonal populations in North America and Europe is that the three archetypal lineages contain linked genes that make these parasites adept at oral transmission and take advantage of the parasite's intermediate host life cycle to expand exclusively asexually (Su et al. 2003). This model is supported by evidence that suggests strains from the three archetypal lineages are relatively poor at outcrossing, tend to favor sexual self-mating in cats, and that a mating type locus may exist on Chromosome Ia (Khan et al. 2006; Fux et al. 2007; Khan, Taylor, et al. 2009). Once a clonal population structure exists, it would limit the opportunities for mixed strain super-infections due to cross-protective immunity, especially between closely related strains (Waldeland and Frenkel 1983; Jensen et al. 2015). Because sexual recombination in the definitive host relies on the ingestion of co-infected intermediate hosts, sexual outcrossing and diversification would be limited by the relative absence of co-infected intermediate hosts in clonally limited populations. However, recent data indicates that mixed genotype co-infections occur with high frequency, at least among wildlife in North America (Gibson et al. 2011). Additionally, systematic sampling of wildlife displayed greater diversity of *Toxoplasma* strains in sea otters off the coast of North America, in marsupials in Australia, and in zoonotic human outbreaks in Africa, Suriname, and Brazil than the previously expected clonality (Yera et al. 2014; Demar et al. 2007; Rajendran, Su, and Dubey 2012; Vaudaux et al. 2010; Parameswaran et al. 2010; Sundar et al. 2008). Furthermore, it is difficult to discern if apparent clonality results from sexual self-mating, unisexual crossing between two genetically

similar strains, or from asexual expansion using only the relatively few markers that have been applied toward genotyping the current population.

An alternative hypothesis recently put forward to explain the clonal structure in North America is that sexual expansion of a limited number of highly fit clones has resulted in the clonal population structure, rather than asexual clonal expansion (Wendte, Gibson, and Grigg 2011; Wendte, Miller, Lambourn, et al. 2010). Support for this model comes from evidence that the three clonal archetypes in North America are in fact highly fit progeny from a recent genetic cross between two discrete ancestral lineages (Grigg, Bonnefoy, et al. 2001). Expansion of this work by Boyle, using EST and genomic sequence data established that the clonal Type I lineage was the product of a cross between a Type II parental strain and an unknown "α" lineage, whereas Type III was a cross between a Type II parent and an unknown "β" lineage (Boyle et al. 2006). The clonal lineages have previously been shown proficient at sexual self-mating. During the sexual cycle, a single cat can produce over 100 million highly infectious oocysts, thus, definitive host infection can massively expand the infection base of a single isolate as has been seen in natural outbreaks whereby large numbers of intermediate hosts are infected by ingesting oocyst contaminated water (Wendte, Miller, Lambourn, et al. 2010).

Sexual recombination between closely related strains, as within clonal lineages, necessitates using whole genome sequencing to resolve whether an isolate is an admixture, or is in whole genome linkage disequilibrium and has expanded asexually as a clone. It has been previously shown that a lack of resolution used to characterize isolates of *Toxoplasma* may lead to the mis-grouping of these isolates with other similar, but not identical, genotypes. An example of biased genomic characterization was found in the strain P89. P89 was originally

categorized as a I/III cross by a select few markers (Sibley et al. 1992); later, after the ancestral lineages of Types I and III were identified as the α and β lineages, P89 was labeled as the ancestral α strain (Boyle et al. 2006). However, when its whole genome was sequenced, P89 was discovered to be a genomic admixture of strains bearing alleles from both the Type I and III lineages, whose genome had undergone significant genetic drift relative to both I and III (Sibley et al. 1992; Minot et al. 2012). Similarly, many of the atypical strains from South America were initially mis-categorized as clonal Type I strains due to the limitations of the genotyping PCR-RFLP markers used to classify them (Ferreira Ade et al. 2004). In North America, Type X isolates were originally classified as Type II strains until polymorphism in a single marker (GRA6) suggested these strains were in fact divergent from the canonical Type II clade (Conrad et al. 2005; Miller et al. 2004). Since, Type X was characterized as the fourth clonal lineage of North America after more genotyping markers were applied (Khan, Taylor, et al. 2009).

Although Type X has been described as the fourth clonal lineage of North America by Khan et al, 2011, work by Sundar suggested that Type X may rather exist as at least 2 recombinant strains based on a single unique PCR-RFLP marker that differentiates two alleles, referred to as A and X, from Type II (Sundar et al. 2008). More recently, two Type X strains (ARI, RAY) were whole genome sequenced, and determined to be genetic recombinants between a Type II line and another lineage, referred to as "γ", different from α and β (Lorenzi et al. 2016). The previous characterization of Type X as the fourth clonal lineage in North America appears to have suffered from a lack of adequate markers to capture the true diversity of Type X genomes. This dissertation has amassed a set of Type X strains to determine whether this clade

exists as a clonal lineage, or a recombinant clade, in addition to determining its exact genetic ancestry.

It is now evident that the limited number of PCR-RFLP markers discussed above fail to capture the true genetic diversity within the *Toxoplasma* population genetic structure (Shwab et al. 2014). With the accessibility of whole genome sequencing it is apparent that the population genetic structure of *Toxoplasma* needed to be re-visited at whole genome sequence resolution. The White Paper, which sequenced 62 *Toxoplasma* genomes to characterize the extent of total genetic diversity was completed during the time of this thesis, established that the majority of *Toxoplasma* strains are recombinants, and parsed the population into six recombinant clades (Lorenzi et al. 2016). Previously, only the three major clonal lineages were available for sequence analysis. Not only did the White paper increase the number of sequenced *Toxoplasma* genomes, but because the strains were selected based on divergence from the known genotypes (established using the standard set of Su markers), this allowed the community to better characterize the number of genetic ancestries within the population, and how they are distributed across the diverse strains sampled. The comparative analysis of the 62 genomes also showed evidence of strong geographic segregation between groups within the population, which supported previous work done using the existing, but limited markers (Minot et al. 2012; Shwab et al. 2014).

Following the increase in whole genome sequencing quantity that the White paper brought to the *Toxoplasma* field, it was necessary to reassess at the genome level the admixture inherent within the population. To determine and visualize the number of genetic ancestries for any given population of Toxoplasma strains, the new genomes from the White paper as well as a few newly sequenced genomes from the Grigg lab were utilized in a collaboration between

the Grigg lab and the Parkinson group of computational biologists at the University of Toronto to develop a new agnostic clustering program termed PopNet. This program was created in conjunction with the work done in this thesis and while it is used to analyze the clonal lineages examined here, PopNet was initially created to analyze diverse populations to allow for a more detailed analysis of genetic admixture within the population structure of *Toxoplasma.* PopNet determines the number of genetic ancestries within a sliding window of discrete genomic blocks (*i.e.,* 1kb, 10kb, or 100kb) and then computes how specific blocks of different ancestry are inherited across chromosomes within the population (Zhang et al. 2017). PopNet then paints each block based on which of the identified ancestries it shares the most common SNPs with, which allows for a highly resolved visualization of recombination blocks along each chromosome for each strain within any given analyzed population. PopNet identifies only common SNPs shared between strains and for the 62 sequenced genomes, it identified 14 distinct ancestries, significantly more than the 6 clades previously identified for the same population by the White paper (Lorenzi et al. 2016). Additionally, PopNet allowed the resolution and painting of blocks of common inheritance across the genomes of these strains that has previously only been displayed as an average across the whole genome rather than painting for each chromosome. The analysis identified highly mosaic genomes among the diverse array of strains thus far sampled and showed that recombination has occurred more frequently than has been previously envisaged. The diversity inherent across the population suggested that admixture diversity may also be inherent within subsets of the *Toxoplasma* population as well as across the population.

In nature, uniparental mating is common, such as in fungi, where it generates clonal genetic population structures, and significant structural variation. This structural variation is

readily distinguished from asexual expansion, which results in the step-wise accumulation of SNP mutations that are unique to a specific strain (Heitman 2010; Roach and Heitman 2014; Sun et al. 2014). A major focus of this thesis has been to look within each of two clonal lineages at whole genome resolution to determine whether these strains are in genome-wide linkage disequilibrium and are being expanded exclusively asexually, as has been proposed previously, or by unisexual or self-mating in the definitive felid host. To assess this, whole genome sequencing was performed to calculate total allelic diversity, differences in somy, CNV, gene dosage, sequence haploblock recombination, and other hallmarks of genome evolution that would occur during uniparental mating.

Minimum Ancestry Estimates and Timelines in the Establishment of the Clonal Lineages

Due to the flexibility of the *Toxoplasma* heteroxenous lifecycle and the fact that the parasite can lay dormant and in geographical isolation within an intermediate host for the life of that host, molecular clock estimates for the origin of certain clones within the population genetic structure are widely varied and depend heavily upon the assumptions used to model its phylogenetic diversity. While it is generally accepted that a bottleneck occurred in the evolution of the current dominant clones that comprise *Toxoplasm*a's population in North America and Europe, the causative factors for it are widely debated. Initial work assumed that sexual recombination was rare across *T. gondii* isolates and examined genomic regions under neutral mutation pressures, such as within introns of housekeeping genes. These neutral regions were utilized to determine the extent of genetic drift among strains within a clonal lineage.

Using these parameters, the Sibley and Ajioka groups estimated that *Toxoplasma*'s clonal lineages were derived from a single genetic cross approximately 10,000 years ago (Su et al. 2003). The same group suggested that the more divergent strains found in South America were derived from similarly infrequent crosses and had undergone independent genetic drift for approximately one million years. They postulated that the recent expansion of the clonal lineages in North America and Europe was due to the ability of these clones to expand exclusively asexually via carnivory among intermediate hosts (Khan et al. 2006; Fux et al. 2007; Su et al. 2003). A more recent study examined 13 global populations totaling 168 strains and suggested that *Toxoplasma* differentiated from its most recent common ancestor (*H. hammondia)* 11 million years ago and then underwent a global population sweep between 150,000-1.5 million years ago. These authors attribute this global sweep to the domestication of felines across the developing agricultural world (Bertranpetit et al. 2016).

All of these studies share a number of key assumption flaws that do not allow for an accurate determination of the molecular clock or time to minimum ancestry. Each study has utilized the mutation rate that was previously calculated for *Plasmodium falciparum* because no mutation rate calculation has been performed for *Toxoplasma.* However, unlike *Toxoplasma, Plasmodium* is an obligate heteroxenous organism which must undergo sexual replication as a mandatory part of every successful transmission, and it lacks the ability to expand asexually between intermediate hosts (Meissner et al. 2007). Hence, its mutational clock is likely quite different from *Toxoplasma's*. Furthermore, *Toxoplasma* has the ability to undergo sexual recombination uniparentally and unisexually to produce clones that are nearly identical to the parents. Thus, it is difficult to distinguish asexual expansion from that of sexual self-mating, especially within highly clonal populations where most strains are highly similar to one another

(Cornelissen and Overdulve 1985). With the discovery of ancient *Toxoplasma* DNA intermixed with ancient human DNA in mummies from Egypt, it may be possible to better elucidate the ancestral lineage of this generalist parasite, however this strategy remains to be tested (Khairat et al. 2013).

Identification of Virulence Determinants for *Toxoplasma*

Multiple factors contribute to the severity of disease attributed to a particular strain of *Toxoplasma* including parasite developmental stage, host immunity, polyparasitism, and especially parasite genotype (Wendte, Gibson, and Grigg 2011). Virulence is one of the most easily assayable phenotypes for parasite fitness within its host. In virulence, host death correlates with maladjustment of the parasite to the host. *Toxoplasma* is best adapted to its host when it can produce a chronic infection without morbidity or mortality that would inhibit host survival and thus parasite spread. Virulence varies depending on the host infected. Most of the work to determine virulence genes in *Toxoplasma* has focused on murine virulence between the clonal populations due to the mouse's usefulness as both a model organism and a definitive part of the *Toxoplasma* lifecycle in nature as well as the lab.

The increase in sequencing of known genomes of *Toxoplasma* has allowed for a more definitive comparison of the *Toxoplasma* genomes to those of related parasites such as *Plasmodium, Neospora,* and *Sarcocystis.* Through this work, it was determined that while the genomes of *Toxoplasma, Neospora,* and *Hammondia* were highly syntenic, *Toxoplasma* contained a set of gene families that were extensively expanded within the species, especially in comparison to these other parasites. These gene families were collectively referred to as

secretory pathogenesis determinants, or SPDs, and occurred as discrete haploblocks of gene arrays that were comprised mainly of the secreted ROP and GRA proteins, and SRS surface proteins thought to be important in parasite entry and modulation of host immunity. Proposed roles for the SRS genes include providing adhesion properties to facilitate attachment and entry into the diverse array of host cells this generalist parasite infects. ROP kinases, and GRA proteins have been identified as highly polymorphic effector proteins that hijack immune signaling pathways and function as virulence factors that enhance *Toxoplasma* pathogenesis in animal infections (Hakimi, Olias, and Sibley 2017). The combination of SPDs inherited is thought to determine host range and specificity (Lorenzi et al. 2016).

Gene	ROP5	ROP16	ROP17	ROP18	ROP38	GRA6	GRA7	GRA15	GRA24	GRA25
Type I										
Type II										
Type III										
Function	Assists ROP18	Prolonged STAT3/6 activation	Inactivates IRGs	Inactivates IRGs	Downregulates MAPK pathway	NFAT4 activation	Irga6 specific inactivation	Increased NFkB	Promotes p38 MAPK activation, IL-12 secretion	Induces CCL2 and CXCL1 in mice macrophages
Interactions	ROP18, GRA7, ROP17		ROP5, ROP18	ROP5, GRA7, ROP17			ROP5, ROP18, ROP2, ROP4	GRA24	GRA15	
Discovery Paper	Fleckenstein 2012	Saeij Nature 2007	Etheridge 2014	Saeij Science 2006	Peixoto 2010	Ma 2013	Hermanns 2015	Rosowski 2011	Braun 2013	Shastri 2014
Other Papers	Niedelman 2012, Behnke 2011, Hermanns 2015	Butcher 2011, Yamamoto 2011	Zhao 2014, Zhang 2014	Behnke 2015, Hermanns 2015, Shwab 2015	Melo 2013, Yang 2013, Fritz 2012		Yang 2015, Dunn 2008, Alaganan 2014	Yang 2013		English 2015

Table 1-1: Virulence determinant genes in *Toxoplasma*
Known virulence determinant genes discussed below are summarized here with red indicating that the strain Type labeled has the capacity listed in the function row. Respective discovery publications as well as key publications that elucidated the function of these genes are listed. Predicted interacting partners are listed as derived from these papers.

Previously, an experimental cross between two mouse avirulent strains (II crossed with III) showed that virulence and infectivity in mice is a multigenic trait and that virulence alleles can be epistatic depending on the parasite genetic background. This epistasis can be removed by genetic recombination leading to hybridization between these haploid organisms (Grigg, Bonnefoy, et al. 2001). Work by the Boothroyd and Sibley groups later determined the gene products responsible for the acute virulence phenotype by using a forward genetic approach. Several of the 39 progeny from the II by III cross of two avirulent parental strains possessed different levels of murine virulence. This range of progeny virulence was utilized in a Qualitative Trait Loci (QTL) mapping of the cross using 250 unique genetic markers distributed across the 14 chromosomes of *Toxoplasma*, where five genomic regions were identified that contained candidate virulence genes (Khan et al. 2005). These regions were narrowed based on expression differences corresponding to the polymorphisms in a singular gene, ROP18 (Saeij et al. 2006; Taylor et al. 2006). Similarly, in a cross between Type I and III parents, 34 progeny were genotyped at 175 QTL markers to map genes contributing to murine virulence differences (Taylor et al. 2006). This secondary cross yielded the same ROP18 gene as the dominant genetic marker of acute murine virulence, but it was likewise encumbered by the same limitations of using only the limited set of genome markers available. Due to the lack of genetic markers, combined with a small number of progeny and subsequently limited genetic crossover events, the genomic regions containing the virulence enhancing genes were quite large. Indeed, a lack of well characterized sexual crosses between *Toxoplasma* types, as well as *Toxoplasma*'s extreme linkage disequilibrium among clonal archetypal lineages has extremely limited the use of genome wide association studies (GWAS) and population-based studies to determine virulence genes among strains (Taylor et al. 2006).

In addition to their individual activity as virulence genes, known combinations of virulence enhancing genes have been described, such as ROP5 and ROP18 (see below) that act cooperatively to overcome host immunological resistance and produce a virulent infection when in the correct allelic combination within the infecting parasite (Behnke et al. 2015; Shwab et al. 2016). It was this combination of a Type III ROP5 allele and a Type I ROP18 allele that resulted in the virulent S23 clone derived from avirulent parents that was originally described by Grigg et al. 2001. However, it is clear that ROP5/ROP18 is not sufficient to fully account for mouse virulence, as there are several highly virulent strains that possess avirulent allelic combinations of ROP5/ROP18, such as BOF and CASTELLS (Khan, Taylor, et al. 2009). These incongruous strains must therefore encode unknown virulence factors or combinations that have the potential for increased transmission, altered host potential, or are necessary for *Toxoplasma* to combat non-murine host defenses. Hence, as part of this thesis, the number of *Toxoplasma* crosses was expanded, and crosses between the more divergent lineages of *Toxoplasma* were specifically targeted, and genetically interrogated with whole genome sequencing, in order to increase marker density and reduce the size of the genomic regions that associate with the mouse virulence phenotype used to identify novel virulence factors.

Toxoplasma Specific Pathogenesis Determinants that Manipulate the Host Environment

Work within the last decade has mapped the factors encoded by *Toxoplasma* that trigger host immunity to regulate parasite proliferation. Specifically, parasitic pathogenic proteins

(such as profilin) trigger the host innate immune response via TLR 11/12 (Gazzinelli et al. 2014; Yarovinsky et al. 2005) or the inflammasome (Gorfu et al. 2014). Detection of the parasite causes a signaling cascade within infected cells, which leads to the production of IL-12 and IL-18, which stimulate the production of IFN-γ in natural killer and T cells, as well as TNF-α from monocytes. These innate factors cause activated host cells to upregulate innate recognition factors such as immunity-related GTPases (IRGs) and Guanylate-binding proteins (GBPs) that activate inflammatory transcriptional factors to protect against acute *Toxoplasma* infections (Hunter and Sibley 2012). Importantly, certain strains of *Toxoplasma* are known to actively manipulate the host cellular environment in a strain-dependent manner to facilitate productive infection. The level of activation of the innate immune system has been shown to be manipulated by the *Toxoplasma* parasite, in a manner that varies between strains (Hakimi, Olias, and Sibley 2017).

Toxoplasma must carefully control its own proliferation and subsequent cellular lysis in order to ensure host survival. Thus, the parasite has developed mechanisms to carefully regulate the immune system of the host in order to produce a productive chronic infection with transmissible cysts that can perpetuate the *Toxoplasma* life cycle. To allow its success in nature and ability to infect such a broad host range, *Toxoplasma* has evolved numerous strategies to antagonize the wide-variety of immune systems across its host range. Consequently, it has evolved numerous species-specific virulence factors to facilitate the immune control that ultimately determines its host range. Although virulence is a highly complex, multi-genic trait that depends on multiple interactions between the host and the parasite, so far, only a handful of well-characterized parasite-protein virulence genes have been identified using crosses performed between the most common clonal lineages: ROP5, ROP16, ROP18, ROP38,

GRA15, GRA24, and SRS29c (Saeij et al. 2006; Taylor et al. 2006; Saeij et al. 2007; Fleckenstein et al. 2012; Wasmuth et al. 2012; Braun et al. 2013; Melo et al. 2013).

The majority of these virulence gene candidates were found via QTL analysis as discussed above. The gene candidates were narrowed based on differences between the clonal lineages that could impact parasite virulence. Primarily, these have been polymorphisms in the genes, gene expression differences between the strains involved, and secretion potential based on protein predictions of the gene candidates. These genes have then traditionally been confirmed via single gene deletions to verify gene activity, and in some cases, mechanism of action.

All previously known virulence factors have an influence on host immunity to *Toxoplasma*, whether by biasing the host toward a productive immune response, driving host immunity away from parasite niche detection within the host cells, or modulating *Toxoplasma's* growth by upregulating the host immune system (Saeij et al. 2006; Taylor et al. 2006; Saeij et al. 2007; Fleckenstein et al. 2012; Hakimi, Olias, and Sibley 2017). Moreover, the majority of these known *Toxoplasma* virulence factors seem to function in combination with another partner protein (often a pseudokinase and kinase working in parallel to produce optimal activity for their strain-specific activity) (Behnke et al. 2012; Fleckenstein et al. 2012; Jensen et al. 2013). The virulence of a strain of *Toxoplasma* relies on both the allele of the gene in question as well as the background genetics for the other interacting virulence factors that are present in the same genome.

In the murine host, *Toxoplasma* utilizes a set of secreted proteins to defend against host intracellular immune proteins in order to protect itself from detection and destruction by the host innate immune response. For example, ROP18 has been shown to protect the

parasitophorous vacuole by phosphorylating host Immunity Related GTPases (or IRGs; especially Irga6) that come into cytosolic contact with the PVM (Hermanns et al. 2015; Steinfeldt et al. 2010). Phosphorylation by ROP18 promotes the release of IRGs from the PVM. This prevents the IRGs from accumulating on the PVM and marking it for lysosomal destruction (Saeij et al. 2006). Further examination determined that only certain alleles of ROP18, specifically the Type I and II strain alleles, are capable of driving this IRG release which promotes virulence in infected mice, but only in certain mouse strains (Behnke et al. 2015; Hermanns et al. 2015; Shwab et al. 2016; Khan, Taylor, et al. 2009). A more recent characterization that investigated parasite protein-protein interactions indicated that ROP18 also complexes with a number of other virulence proteins in a strain-specific manner to alter *Toxoplasma* murine virulence. Both Type I and II strains are known to express virulent alleles of ROP18, but the original QTL that indicated ROP18 as a major virulence factor also contained several other peaks, which were readily identified using a secondary QTL analysis, done by holding ROP18 constant to look for secondary interacting proteins (Saeij et al. 2006). This secondary analysis identified an interaction with the pseudokinase ROP5, as was further verified by investigating knock-outs of either ROP18, ROP5, or both proteins in combination (Niedelman et al. 2012; Etheridge et al. 2014; Behnke et al. 2015; Shwab et al. 2016). ROP5 functions to scaffold IRGs and ROP18 and enhance the kinase activity of ROP18 to inactivate its target host IRGs. In fact, it has been hypothesized that the allele of ROP5 is critical for this interaction, which explains why Type II strains (which have a different suite of the tandemly arrayed ROP5 genes) are avirulent (Behnke et al. 2012; Niedelman et al. 2012). While both Type I and II strains contain the ROP18 virulent allele, only Type I and III ROP5 alleles are capable of coordinating a virulence-enhancing effect for ROP18 in mice. Thus, because only

Type I strains have virulent alleles at both ROP5 and ROP18, Type I strains kill mice at much lower doses than either Type II or III strains (Behnke et al. 2015; Hermanns et al. 2015; Niedelman et al. 2012; Shwab et al. 2016).

The working model is that ROP5 contacts host IRGs that bind to the PVM and prevents their oligomerization prior to interaction with the ROP18 kinase (Fleckenstein et al. 2012). Once the IRG is associated with ROP18, the IRG complex gets phosphorylated and a key effector/detection arm of the innate cellular immune response is inactivated. While it has been established that only a certain combination of ROP5-ROP18 alleles is sufficient for virulence, no further papers have been able to resolve the molecular details of exactly how ROP5 promotes ROP18 phosphorylation to inactivate host IRGs or why the correct combination of these proteins increases strain virulence (Adomako-Ankomah et al. 2014; Behnke et al. 2015; Behnke et al. 2011; Niedelman et al. 2012; Shwab et al. 2016; Steinfeldt et al. 2010; Behnke et al. 2012). Protein-protein interaction studies have suggested that ROP17 also interacts with the ROP5-ROP18 virulence-complex. It has been postulated that ROP17 may interact with host Guanylate Binding Proteins (GBPs) (Etheridge et al. 2014). Indeed, deletion of the Type I allele of ROP17 does further reduce the virulence of the attenuated ROP18 knock-out. Additional work has shown that ROP17 may be more involved with cyst burden in chronic infections rather than acute virulence, but this mechanism of action is unclear (Fox et al. 2016; Zhang et al. 2014; Zhao and Yap 2014). In addition, it has been established by co-purification studies that GRA7, although it does not bind directly to ROP18, facilitates the assembly of the ROP5-ROP18 virulence-complex by interacting with the ancillary ROP5 protein (Hermanns et al. 2015). Further, GRA7 is known to cause ROS activation and NF-kB signaling by binding to IRAK4 and TRAF6, two innate immune factor proteins located in the cytosol of the host cell

that promote NF-kB translocation to the nucleus. Further, deletion of both GRA7 and ROP18 leaves *Toxoplasma* completely avirulent, but how this interaction augments the action of the ROP5-ROP18 virulence-complex is currently not understood (Yang et al. 2015; Alaganan et al. 2014). GRA7 is also known complex with ROP2 and ROP4, so it is conceivable that the ROP5-ROP18 virulence-complex is larger than currently characterized and may be expanded by additional strain-specific knockouts (Dunn et al. 2008).

Humans lack most of the IRGs described in rodents, so it is unclear whether the IRG-interacting ROP5-ROP18 mouse virulence complex is relevant during human infections. However, the ROP18 kinase has been shown to phosphorylate other host protein targets, including the transcription factor ATF6β. In human cells, ATF6β degradation is thought to prevent the presentation of antigens from *Toxoplasma* on host cell surfaces, affecting CD8 T cell recognition of infected cells, which could potentially impact parasite virulence during human infection (Yamamoto et al. 2011). Other virulence factors that have been described include the dense granule protein TgIST, which inhibits STAT1. TgIST traffics from the PVM to the host cell nucleus where it binds to both STAT1 and chromatin-remodeling proteins (Gay et al. 2016). Based on available mRNA data, this protein inhibits STAT1 induction of the IFN-γ response early during host infection and promotes parasite infection competency (Hakimi, Olias, and Sibley 2017; Olias et al. 2016).

The above virulence factors inactivate effector arms of host immunity, but *Toxoplasma* also secretes proteins that activate host immunity. These factors are thought to impact parasite pathogenesis and regulate parasite proliferation to promote a chronic, transmissible infection that facilitates the parasite's propagation to new hosts. ROP16 was the first virulence factor discovered that promotes parasite transmissibility. ROP16 is secreted into the host cytoplasm

during invasion and traffics to the host nucleus where it creates a prolonged activation of STAT3/6 (except in the avirulent Type II strain, where a single point mutation blocks this effect) and consequently alters early IL-12 production kinetics, which in turns regulates active parasite proliferation (Saeij et al. 2007). Additionally, it has been shown that ROP16 activation increases the susceptibility of host macrophages (the preferred cellular host of *Toxoplasma*) to *Toxoplasma*, without activating macrophage pro-inflammatory pathways, which allows increased dissemination of the parasite within the host (Butcher et al. 2011; Yamamoto et al. 2011; Hunter and Sibley 2012; Jensen et al. 2011; Melo et al. 2013). Co-deletion studies have shown that Type II ROP16 functions in concert with Type I GRA15 to stimulate host immunity and increase oral infectivity and cyst burden in mice (Jensen et al. 2013). GRA15 is similarly secreted into the host cytoplasm during infection and traffics to the host nucleus. Once in the host nucleus, certain alleles of GRA15 are known to induce the transcription of the NF-κB pathway (Rosowski et al. 2011). This activation has been previously shown to activate IL-12 which is the main pro-inflammatory response that host cells use to activate early effector immunity against *Toxoplasma* (Yang et al. 2013). In co-immunoprecipitation studies of cell lysates, GRA15 has been shown to interact with GRA24, although the nature of their interactions has yet to be determined (Braun et al. 2013). Like ROP18, GRA15 is hypothesized to function as a scaffolding protein, in complex with GRA24 (Braun et al. 2013). Like GRA15, GRA24 traffics from the host cytosol into the host nucleus where it activates the transcription of a number of MAPK genes that ultimately leads to the upregulation of IL-12, as seen previously with GRA15 (Pellegrini et al. 2017). Unlike GRA15, GRA24 directly interacts with histone-modifying enzymes that manipulate the host chromatin to facilitate pro-inflammatory cytokine transcription (Braun et al. 2013). GRA24 has also been shown to affect the levels of

Cxcl1 and Ccl2 in host cells (Shastri et al. 2014). Similarly, GRA6 is another parasite polymorphic effector protein that traffics to the nucleus and activates NFAT4 in an allele-specific manner. Like GRA24, Type I and III strain alleles of GRA6 upregulate the expression of Cxcl2 and Ccl2 which recruit immune cells to sites of active infection. The activation of NFAT4 has been suggested to modulate both the growth of the parasite within host cells (by activating innate immune cells to limit their replication) and also to broadly disseminate *Toxoplasma* in highly mobile, infected immune cells (Ma et al. 2014). Further evidence for this GRA6 immune regulation has been found in exotic strains where complementation of deletions in GRA6 can restore virulence (Fazaeli et al. 2000). In contrast, ROP38 is a rhoptry protein that downregulates the expression of MAPK pathway genes (Peixoto et al. 2010). In those strains that highly express ROP38, NF-κB activation is lowered and results in lower levels of anti-parasitic IL-12 (Melo et al. 2013; Fritz et al. 2012). A small number of virulence factors are now being discovered that regulate the immune response to *Toxoplasma* without being secreted into the host cell. One of these is the surface expressed antigen SRS29C (also annotated as SRS2) that is a negative regulator of virulence controlled solely by strain-specific expression levels, rather than allelic type (Wasmuth et al. 2012). When alleles of Type I SRS29C were upregulated to Type II protein expression levels, parasites displayed both delayed dissemination as well as decreased virulence, indicating that even surface proteins may play a role in the manipulation of acute disease within infected hosts.

As different strains of *Toxoplasma* evolve to subvert host immunity across the parasite's broad host range, a parallel adaptation within the host immune system is occurring to detect and defend against *Toxoplasma*. This co-evolutionary arms race is potentially driving the evolution of *Toxoplasma* genetic diversity and its ability to extend its host range (Gazzinelli et al. 2014;

Hunter and Sibley 2012). While it is largely thought that *Toxoplasma* is capable of producing chronic infections in all animal hosts, the mechanism of action to produce these chronic infections likely varies with the infected host species. Studies in rats have identified a single rat gene, the inflammasome sensor Nlrp1b, that confers sterile immunity and complete resistance to all tested strains of *Toxoplasma* (Cavailles et al. 2006; Cirelli et al. 2014; Cavailles et al. 2014). Additionally, recent work in human cell lines suggests that humans combat *Toxoplasma* infection differently than rodents. The differences between these host immune systems is such that studies performed in mice may not necessarily predict the outcome of human infection. This is largely because mice rely on their diverse range of IRGs and GBPs to intracellularly detect and control acute *Toxoplasma* infection and dissemination. Humans entirely lack functional IRGs, and GBPs in human hosts do not appear to confer protection in a similar way that mouse GBPs control parasite infection (Dupont and Hunter 2012; Gazzinelli et al. 2014; Hunn et al. 2011; Hunter and Sibley 2012; Selleck et al. 2013; Steinfeldt et al. 2010). This suggests that within the suites of polymorphic effector proteins that *Toxoplasma* encodes, a subset of these effector proteins may confer protection from host immunity in, for example rodents, whereas a different subset is required for successful infection in other intermediate host taxa, such as humans.

Evolutionary selection can function at the parasite lifecycle stage as well as on the strain. Two sexual stage-specific *Toxoplasma* genes, AAH1 and AAH2 encode for aromatic amino acid hydroxylases that are required for successful completion of the parasites' sexual cycle in the definitive feline host. Without these genes, feline infections yield substantially reduced numbers of oocysts, however these genes do not appear to have a dramatic effect on murine cyst burden (Wang, Verma, et al. 2017). This interplay of diversity between the parasite

and its host is likely a major driving factor in the evolution of *Toxoplasma's* diverse population genetic structure. In fact, differences in the activation of host immunity across the parasite's broad intermediate host range may ultimately determine which parasite strains expand across, or are the best fit for, any given animal species to produce a transmissible infection. This co-evolution of *Toxoplasma* and its hosts makes sense for a highly successful generalist parasite as host selection determines which strains are expanded as transmissible infections across *Toxoplasma's* diverse genetic population structure.

The Role of Sexual Recombination in the Population Dynamics of *Toxoplasma* Evolution

Toxoplasma is a highly successful parasite with the ability to infect any warm-blooded animal on the planet as well as an incredibly robust sexual cycle with the ability to both expand the current population by uniparental replication and diversity by mixed strain co-infection. Recombination between even avirulent strains of *Toxoplasma* has previously been shown to alter this parasite's biological potential by bringing alleles from different genetic backgrounds into varying combinations within the same genomes (Grigg, Bonnefoy, et al. 2001). Indeed, diverse host selection of the admixture progeny that result from sexual recombination may be the key evolutionary driving factor underlying *Toxoplasma* evolution and host adaptation as intermediate hosts select for parasites that cause chronic avirulent infections (Agrawal 2006). Despite its fecund sexual cycle and due mainly to its supposedly clonal population structure, propagation of *Toxoplasma* has been postulated to occur mainly via asexual replication, while genetic recombination via meiotic sexual replication is thought to be infrequent due to a lack of

admixture observed in the population as a whole (Wendte, Miller, Lambourn, et al. 2010; Sibley and Ajioka 2008; Grigg and Sundar 2009; Boothroyd 2009; Minot et al. 2012).

However, the studies that hypothesized that *Toxoplasma* sexual recombination is rare were hampered by a number of crucial limitations. One limitation to these conclusions is that most population studies have utilized relatively limited numbers and resolutions of genetic markers which were designed to differentiate only the largest divergences in these parasites' genomes (Minot et al. 2012; Su et al. 2012). For instance, low marker resolution initially caused Type X strains to be classified as Type II strains despite their genomic diversity at several key markers (Dubey et al. 2011; Khan, Dubey, et al. 2011; Sundar et al. 2008). Additionally, most studies have focused on the diversity across the entire population and ignored any variation that has occurred within the clonal lineages, even when phenotypic differences were observed (Parameswaran et al. 2010; Khan, Behnke, et al. 2009; Yang et al. 2013; Verma et al. 2015). Due to this bias toward most dissimilar strain interrogations, lower diversity levels, such as those that derive over shorter time scales, have not been interrogated within the *Toxoplasma* population. Because of the limited diversity within the parasite population, it is possible that sexual recombination can be masked when strains of similar ancestries recombine, as can happen when the definitive host is infected by ingesting an intermediate host that carries a co-infection of genetically similar strains. Cryptic uniparental and unisexual meiotic recombination likely masks the extent to which sexual recombination is occurring within the population.

However, with the advent and increased utilization of whole genome sequencing in the *Toxoplasma* field, an increased appreciation of diversity within the genetic population structure is becoming apparent. It was a major goal of this thesis to interrogate two common, clonal clades of strains within *Toxoplasma* population genetics to determine the extent to which these strains

are in linkage disequilibrium and are expanding asexually versus undergoing sexual recombination. Additionally, the new availability of the maternally inherited apicoplast genome can be utilized to compare the nuclear genome to maternally inherited genomes which do not undergo meiotic recombination. Incongruence between the inheritance of nuclear versus organellar genomes within a parasite is a hallmark of sexual replication for the parasite. This thesis aims to test the sufficiency of the clonal theory of expansion for describing the population structure of *Toxoplasma.* To assess this, whole genome sequencing was performed to calculate total allelic diversity, differences in somy, CNV, gene dosage, sequence haploblock recombination, and other hallmarks of genome evolution that would occur by uniparental mating. While whole genome sequencing is useful for determining broad recombination patterns within a sexual clade, genomes must be interrogated in sequence haploblocks in order to determine whether diversity stems from genetic drift, which is derived from mitotic replication errors and will be evenly distributed across the genome, or from sexual recombination, which distributes genetic diversity in haploblocks across the genomes, which can be clearly delineated by recombination crossovers.

A collection of strains belonging to two clonal lineages, Type II and Type X, were sequenced at whole genome resolution to infer whether they replicate in nature predominantly asexually, or sexually by uniparental mating (which is virtually indistinguishable from asexual recombination by low-resolution marker typing), to identify the most likely route of transmission for these highly successful clades. SNP diversity was quantified in sliding windows, typically in 1-10kb sequence blocks, and haplotypes were established. This analysis established that recombination was occurring at high frequency within these clades of supposedly clonal strains. While the Type II strains were closely related to one another, SNP diversity within this clade was

geographically clustered and was sufficient to readily detect sequence haploblocks of varied ancestry between these strains. Several of the North American Type II isolates were in fact mosaic clones containing haploblocks of Type II sequence that were indigenous to either Europe or North America. Hence, Type II exists as a clade of highly similar strains that preferentially recombine unisexually but which appear to be expanding asexually when only a few low-resolution markers are used. Whereas for Type X, increased marker density and WGS analysis established that this clonal lineage was not clonal, but rather resolved as a sexual clade of recombinant strains comprised of at least three distinct ancestries (Type II, γ, and δ) that had recombined across the genomes of the strains within Type X. Both intra and extra clade sexual recombination was not only occurring more frequently than had been previously postulated across the *Toxoplasma* population, but unisexual crossing was likely contributing to the successful expansion of the Type II lineage in nature.

During the course of this work, the Type X lineage was resolved into 12 distinct haplotypes that appeared to exist as a sexual clade of natural recombinants resembling F1 progeny from a genetic admixture between a Type II strain and a novel genotype, γ/δ. Interestingly, the haplotype that expanded to cause the majority of avirulent infections in sea otters was highly virulent to mice. Our data supported a virulence shift model whereby generalist pathogens like *Toxoplasma* utilize their sexual cycles to produce virulent strains that can be maintained cryptically in nature according to their differential capacity to cause disease across the pathogen's broad intermediate host range. This type of "zoonotic selection" has important public health implications. Strains capable of causing fatal infections can persist in nature by circulating as chronic infections in tolerant intermediate host species that act as reservoirs for potential epidemic disease. Because all Type X strains possessed avirulent combinations of

ROP5 and ROP18 alleles, the opportunity to identify new murine virulence alleles was realized. A forward genetic approach was pursued to separately cross two virulent Type X clones with the avirulent Type II parent to identify novel virulence alleles, as well as to increase the number of recombinant progeny available to the field for other quantitative trait studies. Furthermore, this work aimed to deduce the amount of genomic recombination crossovers that occurs within a sexual cross by whole genome sequencing individually isolated sexual progeny in a novel manner eliminating the bias that was inherent in previous isolation methods. This work was also done to more accurately enumerate the extent to which outcrossing versus self-mating occurs during sexual replication. In fact, work shown here determined that sexual recombination is occurring in a more biased manner within the definitive host than was previously envisaged and established that sex was biased toward similar strain mating, which may explain clonal population structures for pathogens, such as *Toxoplasma,* that possess highly fecund sexual cycles.

Chapter 2 - Uniparental Mating and Sexual Expansion of the *Toxoplasma* Type II Clonal Lineage

Kennard, Andrea[1]; Zhang, Javi[2]; Shen, Kui[3]; Quiñones, Mariam[3]; and Grigg, Michael E.[1]

[1]Laboratory of Parasitic Diseases, NIAID, National Institutes of Health, Bethesda, MD. [2]Department of Biochemistry, University of Toronto, Toronto, ON, Canada [3]Bioinformatics and Computational Biosciences Branch, NIAID, National Institutes of Health, Bethesda, MD.

Introduction

While it is well known that sexual replication can recombine the genomes of the generalist parasite *Toxoplasma,* how this diversity is potentiated across *Toxoplasma's* wide host range and to what extent meiotic recombination shapes the population structure of this haploid protozoan parasite remains unclear. Sexual recombination increases the biological and genetic population diversity of eukaryotes by bringing alleles into novel combinations in their recombined genomes (Boyle et al. 2006; Grigg, Bonnefoy, et al. 2001). Many eukaryotes utilize genetic hybridization to increase their biological diversity: *S. cerevisiae* utilize sex to create beneficial aneuploidy (Selmecki et al. 2015), *Leishmania* sexually expand to increase copy number variation and polyploidy across the genome (Romano et al. 2014), *Cryptococcus neoformans* recombine chromosomal-encoded pathogenicity islands through sex (Sun et al. 2014), and *Plasmodium* recombines VAR genes to alter antigenic presentation (Claessens et al. 2014). Sexual recombination is known to increase the potential for adaptation to the parasites' host environment. What is less clear, is how important sexual recombination is in shaping the

Toxoplasma population structure as a whole, as the species is thought to maintain large populations of clonal types by exclusive asexual expansion (Su et al. 2012).

Toxoplasma can orally infect any warm-blooded animal giving it a flexible transmission cycle across a vast infectious host range. The ability of this parasite to expand asexually by carnivory across its intermediate host range is thought to explain its clonal population structure in North America and Europe. However, *Toxoplasma* possesses a highly fecund sexual cycle when it infects its definitive feline hosts, with yields often in excess of 100 million highly infectious and environmentally stable oocysts per feline (Dubey and Frenkel 1976). During feline infection, parasites either outcross or self-mate. Genetic hybridization between two strains is known to occur readily when a cat is co-infected with multiple strains simultaneously, but in the absence of a mixed strain co-infection, single strains of *Toxoplasma* are known to self-mate as both micro and macro-gametocytes can be produced by a single clone and fertilize each other to produce viable oocysts (Cornelissen and Overdulve 1985; Pfefferkorn, Pfefferkorn, and Colby 1977). Due to its haploid nature, progeny from self-mating are virtually indistinguishable from strains that have been expanded asexually when analyzed using low-resolution PCR-RFLP. Hence, self-mating is thought to sexually expand successful clones, while the degree to which this mechanism affects *Toxoplasma's* clonal population structure remains enigmatic.

When two *Toxoplasma* strains simultaneously infect a feline, genetic hybridization can occur between the strains to create progeny with diversified biological potential from those of the two parental strains (Grigg, Bonnefoy, et al. 2001). Indeed, genetic hybridization between two mouse avirulent clones previously yielded progeny that were dramatically more virulent than either of the two haploid parents (Grigg, Bonnefoy, et al. 2001; Sibley and Ajioka 2008). What is less clear is the extent to which successful strains are being expanded in nature. One model

suggests that these clones are being maintained exclusively asexually and have expanded because they possess increased oral infectivity (Su et al. 2003). Alternatively, sexual self-mating expansion of a single clone has been proposed to explain the parasite's clonal population structure in some geographic niches (Wendte, Miller, Lambourn, et al. 2010). The ability of *Toxoplasma* to expand both asexually or through sexual recombination, makes *Toxoplasma* an ideal species to study how the interplay of sexual and asexual replication shape a eukaryotic pathogen's population structure.

Despite the prevalence of sexual recombination among eukaryotic parasites, clonal populations are abundant. Clonal populations are common for eukaryotic pathogens, but the mechanisms by which they are generated and maintained vary widely. *Plasmodium* clonal populations generally result from drug resistant sweeps (Miotto et al. 2013), *Cryptosporidium* remains clonal due to a strict host adaptation which prevents cross-strain interactions (Awad-El-Kariem 1999), and *Cryptococcus* remain clonal due to a lack of mating type complementation (Billmyre et al. 2014). However, none of these adaptations are likely to explain clonality in *Toxoplasma. Toxoplasma* drug treatments do not select for the drug resistance sweeps evidenced in *Plasmodium* as available drugs have not been systematically applied, nor do they provide a sterile cure following treatment. Additionally, all tested *Toxoplasma* strains are capable of feline infection thus, a lack of interaction in the definitive host is unlikely to inhibit sexual recombination. One theory postulates that a shared inheritance of common Type II haploblocks allowed for oral transmission (and asexual expansion) of the clonal strains, but this adaptation does not restrict sexual replication of these strains (Su et al. 2003; Khan et al. 2006). Furthermore, *Toxoplasma* lacks mating type loci which would influence mate selection, and the majority of sexual crossing experiments reported to date have yielded recombinant progeny. The

lack of impediments to sexual outcrossing in *Toxoplasma* brings into question how often sexual recombination occurs and how it affects the population structure.

South American *Toxoplasma* strains have a population structure expected from a parasite with a definitive sexual cycle, possessing highly polymorphic genetic diversity, extensive recombination of this genetic diversity across genomes, and a wide range of virulence phenotypes in a variety of infected hosts. In contrast, despite a fecund sexual cycle, *Toxoplasma* has a uniquely clonal population structure in both North America and Europe, where the population is dominated by only four clonal lineages: I, II, III, and X (or HG12) (Lorenzi et al. 2016; Zhang et al. 2017; Khan, Dubey, et al. 2011). The predominance of clonal lineages in both North America and Europe, has led to the clonal theory of parasite expansion, which posits that the Type I, II, III, and X clonal lineages are expanding exclusively by asexual replication (Sibley and Ajioka 2008; Tibayrenc and Ayala 1991, 2014). However, the clonal theory of expansion does not consider how this clonality is generated and as asexual replication is indistinguishable from uniparental self-mating, both may contribute to the clonal theory of expansion (Wendte, Miller, Lambourn, et al. 2010).

No study to date has systematically examined the contribution of the sexual cycle in the maintenance of *Toxoplasma's* clonal population structure, as the majority of studies have focused on charting genetic diversity worldwide and did not examine the diversity that exists within a clonal lineage at whole genome resolution (Lorenzi et al. 2016; Sibley et al. 2009; Su et al. 2012). Additionally, a significant portion of the genomes in three of the clonal lineages was inherited by genetic hybridization with the Type II lineage, making the Type II genomic contribution to the population structure more prominent than other lineages (Boyle et al. 2006). Further this shared inheritance within the clonal lineages suggests that Type II genomic

inheritance is beneficial for clonal expansion (Khan et al. 2006). Because of the increased density of genotyping markers, especially microsatellite markers, it is evident that genetic diversity does exist within the clonal lineages, especially the widely prevalent Type II clade. Microsatellites differences, the result of polymerase error during mitotic DNA replication, can be used to differentiate closely-related strains, such as those from recent outbreaks, to determine strain diversity on a limited scale (Demar et al. 2007; Wendte, Miller, Lambourn, et al. 2010; Grigg and Sundar 2009; Vaudaux et al. 2010; VanWormer et al. 2014). Microsatellite typing analyses have demonstrated limited genetic diversity within Type II strains isolated from chickens in Israel and Australia, as well as in Type II-like strains derived from Australian wildlife infections (Verma et al. 2015; Parameswaran et al. 2010; Silva et al. 2017). Further, French patients seropositive for Type II infections displayed a range of disease presentations (Ajzenberg et al. 2009). This genetic and phenotypic diversity observed within a single worldwide clonal lineage makes Type II an excellent clade for the interrogation of the sufficiency of the clonal asexual expansion model to explain the *Toxoplasma* population structure.

The nature of microsatellites, which allows their use for closely-related differentiation of strains, makes them less useful for the interrogation of more geographically diverse strains. However, with the availability of whole genome sequencing, these Type II genomes were interrogated in greater depth to elucidate the diversity between Type II strains identified by the Su marker PCR-RFLP analysis. Work done here with a globally-derived subset of Type II strains has allowed for the novel interrogation of the diversity within a clonal lineage and the clonal theory of expansion by whole genome sequencing. Seventeen Type II strains from across the globe were sequenced using a wide selection of linked and unlinked markers across the nuclear

and organellar genomes of these isolates. Each of these strains were also whole genome sequenced and reference mapped to both the nuclear genome and the maternally inherited organellar apicoplast genome to interrogate genomic evolution and test the sufficiency of the PCR-RFLP analysis to accurately predict the Type II designation. The appreciation of the polymorphisms across sequence typing markers, combined with characterization of unique genomic haploblocks of Type II sequence diversity via SNP density and whole genome interrelationship studies, such as NeighborNet and PopNet, established that strains belonging to the Type II clade display more genetic diversity than was previously envisaged. Importantly, this diversity was not evenly distributed across each Type II genome but was rather localized into distinct haploblocks that were inherited independently across the Type II strains, suggesting they were derived by intra-clade genetic hybridization. From these analyses, both sexual recombination and mitotically acquired genetic drift were observed in these genetic mosaics within the Type II lineage. While low-resolution PCR-RFLP analyses indicated that all strains belonged to the clonal Type II clade, sequencing of the PCR markers and whole genome sequencing of each strain established that sexual recombination between closely-related Type II strains is occurring and that sexual expansion by uniparental mating better explains how *Toxoplasma* is maintaining its clonal population structure.

Results

Type II Strains Comprise a Single Clade of Strains by Low-Resolution Typing

Previous interrogations of Type II strains identified genetic diversity within the clonal lineage by microsatellite typing (Silva et al. 2017; Verma et al. 2015; Parameswaran et al. 2010).

However, this diversity has never been investigated using single-copy polymorphic genes, let alone at the whole genome level, which captures genetic evolution over larger time scales. To examine the genetic diversity inherent within the Type II clonal lineage, 17 previously genotyped Type II strains from across the globe were selected as representatives of the Type II lineage. These strains were selected from 10 locations across 6 continents to encompass worldwide diversity within this clonal lineage (Supplemental Figure 1). To make relative genetic diversity comparisons, 3 Type I, 2 Type III, and 5 Type X strains, all of which have previously introgressed with the Type II lineage, were included in the genome-wide SNP diversity plots (Boyle et al. 2006; Lorenzi et al. 2016). Eight outgroup strains were also included to examine total genetic diversity across the entire *Toxoplasma* population.

Traditionally, ten PCR-RFLP markers have been used to characterize newly isolated strains (Shwab et al. 2014). All of the Type II strains selected for this study were categorized by PCR-RFLP as Type II clonal strains (Figure 1A). These Type II strains, that were colored according to their geographic origin of isolation, resided on a single, well supported branch with no observable diversity within the concatenated-marker maximum likelihood phylogenetic tree (Figure 1B). Similarly, the Type I, III, and X clonal lineages grouped with their respective clonal clades. The grouping of these strains by PCR-RFLP supports their characterization as clonal lineages, however, it is known that the APICO marker breaks the Type II lineage into two separate genotypes (ToxoDB #1 and #3) but, these two genotypes are commonly referred to as the Type II lineage, as all nuclear-encoded markers are in linkage disequilibrium. Included in this analysis were, 4 Type II strains with the ToxoDB genotype #1 designation and 11 Type II strains with the ToxoDB genotype #3 designation.

Greater Marker Resolution Differentiates Type II Strains

A lack of genomic resolution has previously been the culprit of incorrect strain classification, for example when the Type X strains were classified as part of the Type II lineage (Khan, Dubey, et al. 2011; Verma et al. 2015; Silva et al. 2017). To better resolve the Type II clade, the PCR-RFLP markers were sequenced and additional linked and unlinked sequenced markers were added across the genome. The full complement of markers included 10 markers defined by Su *et al.* (Shwab et al. 2014; Su, Zhang, and Dubey 2006) as well as 4 markers previously described by Grigg *et al.* (Wendte, Miller, Lambourn, et al. 2010; Grigg and Sundar 2009; Parameswaran et al. 2010), increasing the sequence used for analysis by 120% (Supplemental Table 1). The ToxoDB PCR-RFLP markers covered 7135 bp of genomic sequence, but only accounted for 18 SNPs across the loci interrogated via RFLP digestion, whereas the sequenced markers covered 15782 bp of sequence and increased the number of SNP positions to 413. Microsatellite markers were excluded from our analyses because they diverge at a much faster rate than traditional typing markers and have the potential to back mutate, so they are less informative for determining ancestry. Microsatellites are more appropriate to interrogate strain differences at a single location or during an outbreak situation, whereas this study aims to determine genomic diversity over longer time scales within a clonal lineage.

To establish the haplotypes for each of the Type II strains, all 14 pan-genomic markers were analyzed using 1000 bootstrap supported maximum likelihood trees for each marker separately and for all markers in a concatenated sequence (Figure 1C). Phylogenetic divisions of 60% or greater bootstrap between tree branches identified unique alleles of independent ancestry. This threshold was chosen because it differentiates the Type II and closely related X strain phylogenies and is thus suitable to differentiate mitotic drift from unique genetic diversity

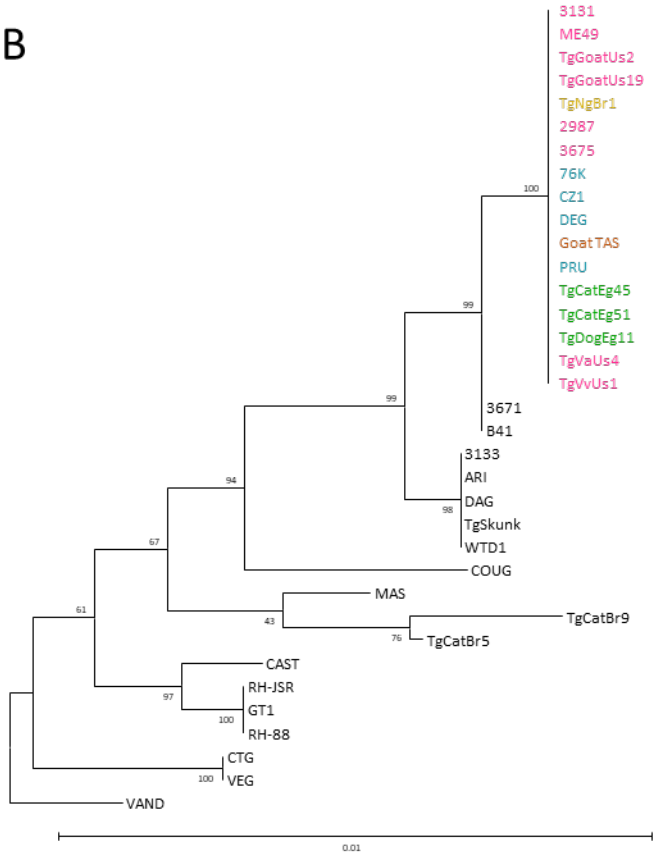
between clades. Analysis of the concatenated tree demonstrated that while Type II remained closely grouped, Type II strains divided into five sub-groups with >60% bootstrap support within the clade (Figure 1C). Of the 17 Type II strains analyzed, only ME49 remained in the original group. While minor mitotic drift is expected within closely-related clades of strains, the number of groups obtained from these sequenced markers was unexpected (Khan, Behnke, et al. 2009; Verma et al. 2015). Because the sequencing performed was limited, accounting for 0.02% of the genomes, these results suggested that whole genome sequencing was necessary in order to fully characterize the diversity inherent within the Type II lineage.

Figure 2-1: Genotyping analyses identify genetic diversity within the Toxoplasma Type II clonal lineage
A) Characterization of 10 standard *Toxoplasma* PCR-RFLP markers by restriction digest across 17 strains of Type II, 3 Type I, 2 Type III, 5 Type X, and 8 outgroup strains. Markers are shown with their chromosomal location, length of sequence, and RFLP digestion enzyme used. Strains are grouped by PCR-RFLP designation and their ToxoDB genotype is shown. Identity of the markers across these strains is listed based on observed digestion compared to known digestion patterns. **B)** Simulated maximum likelihood tree based on allele present using the 10 PCR-RFLP markers. Allele identity from digests were concatenated and used to simulate a maximum likelihood tree with 1000 bootstrap support. All Type II isolates were color-coded based on the geographic location of the isolate. (North America, South America, Eurasia, Africa, Australia). This color scheme is used throughout the figures. **C)** Four additional markers were added to the ten original markers above (markers listed in Supplemental Table 1). All 14 markers were DNA sequenced and concatenated to produce a maximum likelihood tree with 1000 bootstrap support. The tree was based on 15782 bp of sequence across the genomes of these strains. Strains were colored as in 1B (North America, South America, Eurasia, Africa, Australia).

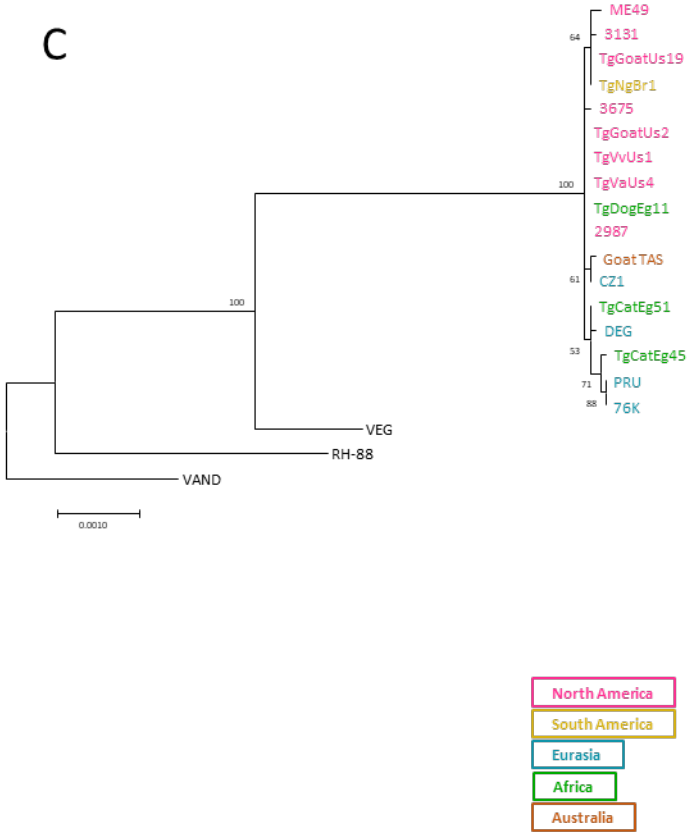
A

Strains	ToxoDB Group	APICO	c228	c292	L358	PK1	SAG1	SAG2	Beta Tubulin	GRA6	SAG3	Strains
RFLP		AflII	BsmAI	HpyCH4IV	HaeIII	AvaI	Sau96I	MboI	BsiEI	MseI	NciI	
		DdeI	MboII	RsaI	NlaIII	RsaI	HaeII	HhaI	TaqI			
Chromosome		apicoplast	Ib	III	V	VI	VIII	VIII	IX	X	XII	
Size (bp)		848	657	689	690	1029	482	1056	527	846	311	
Type I	#10	I	I	I	I	I	I	I	I	I	I	GT1, RH-88, RH-JSR
Type II	#1	II	II	II	II	II	II/III	II	II	II	II	ME49, TgGoatUs2, TgGoatUs19, TgNgBr1
Type II	#3	I	II	II	II	II	II/III	II	II	II	II	2987, 76K, CZ1, DEG, Goat_TAS, PRU, TgCatEg45, TgCatEg51, TgDogEg11, TgVaUs4, TgVvUs1
Type III	#2	III	III	III	III	III	II/III	III	III	III	III	CTG, VEG
Type X	#5	I	II	II	I	II	u-1	II	II	II	II	3133, ARI, DAG, TgSkunk, WTD-1
	#4	I	II	II	I	II	II/III	II	II	II	II	3671, B41
	#66	I	II		I	u-1	I	u-1	II	II	III	COUG
	#28	III	II		III	I	I	I	I	I	I	CAST
	#42	I	I	I	I	u-2	I	I	III	II	III	TgCatBr9
	#19	I	I	I	u-1	u-2	I	III	III	III	III	TgCatBr5
	#17	I	u-1	I	u-1	III	u-1	I	III	III	III	MAS
	#60	I	III	I	III	III	u-2	I	III	III	I	VAND

B



C



Whole Genome Sequencing Reveals the Nuclear-Organellar Incongruence and Reticulation Inherent in the Type II Lineage

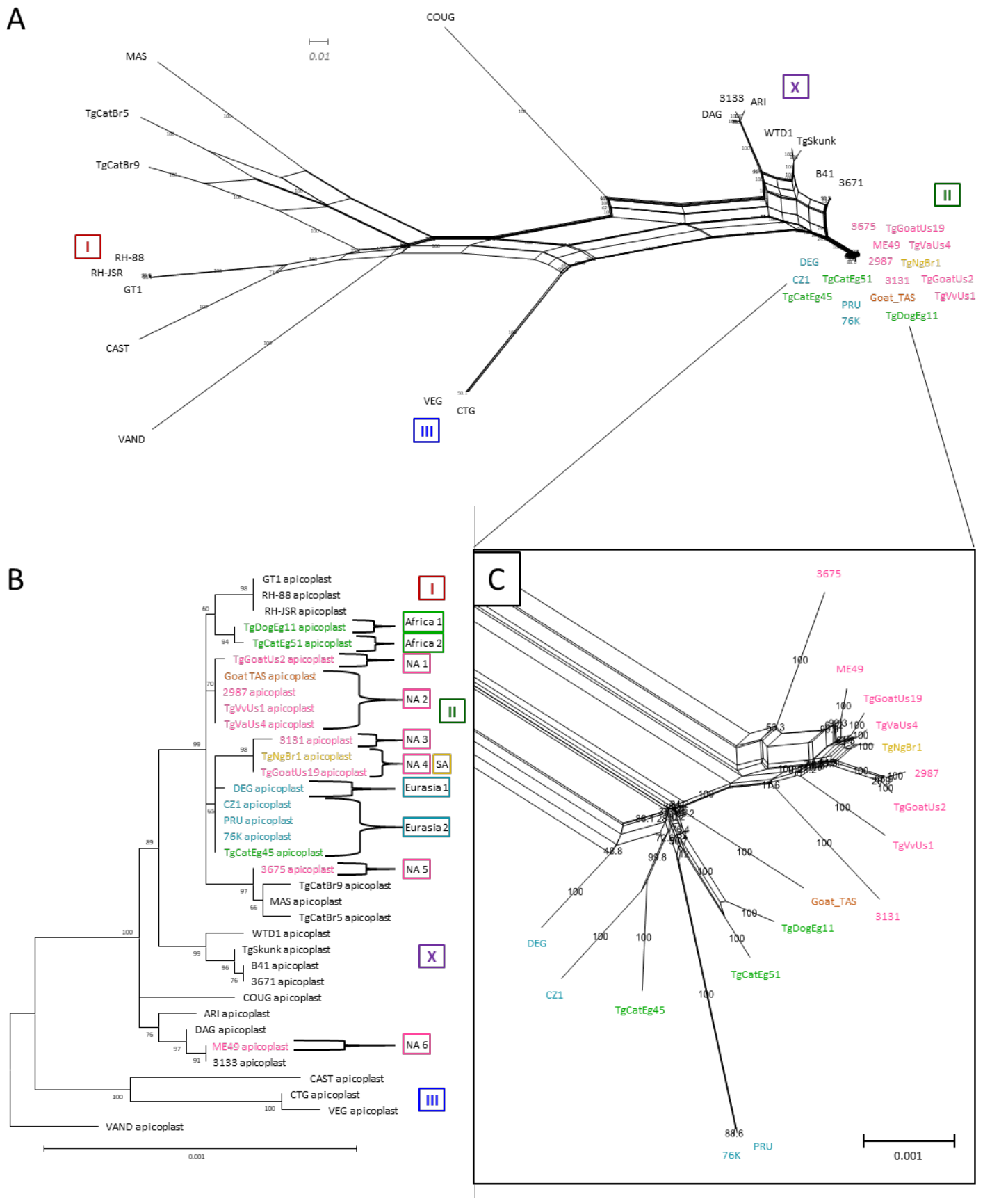
To better resolve the true diversity and genetic history within Type II, whole genome sequencing was performed for each isolate and reference mapped to the canonical Type II, ME49 genome (ToxoDB Version 8.2) and apicoplast organellar genome (GenBank KE138841). As the apicoplast does not undergo meiotic recombination and is maternally inherited, it is ideal for characterizing incongruence between the nuclear versus apicoplast tree topologies that occurs when meiotic recombination shuffles the alleles of nuclear-encoded genes into novel combinations (Ferguson et al. 2005). The 35 kbp apicoplast of each strain was analyzed by a maximum likelihood tree to compare against the nuclear genome (Figure 2B). Mitotic drift within the apicoplast genome was expected due to the age of the Type II lineage (Boyle et al. 2006). Assuming strains are expanding largely asexually, then the accumulation of SNPs during asexual replication should produce trees that are congruent between nuclear-encoded genes and those of the apicoplast. Additionally, strains from the same geographic isolation location would share the majority of the polymorphisms across the genome and should group closely (Silva et al. 2017). While a significant portion of the Type II strains do geographically group, nuclear-organellar incongruence was observed in multiple Type II strains. The Australian GoatTAS and South American TgNgBr1 strains grouped with the North American strains in NA 2 and 4 respectively despite their distant geographical isolation. Additionally, the African *Toxoplasma* strain, TgCatEg45 grouped with the Eurasian strains rather than with the other two African derived strains (TgDogEg11 and TgCatEg51). Ten apicoplast-based groups of strains were identified from this sequencing. Importantly, the ten apicoplast groups displayed a marked incongruence from the ten nuclear inheritance groups based on the extended pan-genomic

marker sequences (Figure 2B and Figure 1C respectively). The incongruence between the nuclear and organellar genomes indicated that sexual recombination is occurring within the Type II clade.

The incongruence observed between the maternally inherited genome and nuclear markers indicated that WGS is required to infer an accurate genetic history model and to interrogate the clonal theory of expansion in the closely-related Type II strains. 17 Type II, 3 Type I, 2 Type III, 5 Type X, and 8 outgroup strains were sequenced and mapped against the canonical ME49 Type II strain to analyze intra-clade diversity. Over 973,800 SNP positions were identified across the genomes and used to assemble a NeighborNet tree based on the divergence between strains (Figure 2A). Indels were excluded due to their poor integration with phylogenetic analysis software. As expected, the Type I and III clonal lineages shared edge blocks of reticulation due to their common ancestry with the Type II lineage, but otherwise grouped as single branches (Boyle et al. 2006). Type X strains grouped onto a number of branches, providing the first evidence that Type X is not a clonal lineage (see Chapter 3) but rather recombinant progeny from at least one sexual cross with Type II (Khan, Dubey, et al. 2011; Sundar et al. 2008). All Type II strains grouped together on a single reticulated branch, as expected for a closely-related clade of strains, whose branch length is consistent with Type II ancestral divergence from the other clonal lineages (Khan, Dubey, et al. 2011; Lorenzi et al. 2016; Shwab et al. 2014; Sibley and Ajioka 2008; Zhang et al. 2017). Close inspection of the Type II branch (Figure 2C), however, showed evidence of both minor mitotic drift within the Type II strains, evidenced by the short branch lengths between strains, as well as reticulation between the strains, such as that found between TgNgBr1 and ME49, indicative of sexual recombination between the strains. As with the apicoplasts (Figure 2B) the nuclear genomes of

the Type II strains generally grouped based on geographic isolation (Figure 2C), with a few notable exceptions such as GoatTAS, a strain isolated in Australia that grouped with the African strains. The strain reticulation combined with the incongruence between the nuclear and organellar genomes tree topologies suggest that Type II strains are recombining their genomes within this closely-related clade. As NeighborNet trees examine the polymorphic differences between strains but ignore the position of these polymorphisms, it is necessary to examine the position of these SNPs across the genomes to determine the evolutionary inheritance of these strains. Positional analysis via SNP density plots is necessary to assess whether these SNPs are inherited in recombination blocks (referred to as haploblocks) indicative of sexual recombination or at random across the genome, as would occur in asexually expanding strains that accumulate mutations from polymerase error.

Figure 2-2: Incongruence between Type II nuclear and organellar genomes identifies sexual recombination within the Type II lineage
A) All strains were whole genome sequenced and reference mapped to ME49. Over 973,800 SNP variant positions were called and used to construct a NeighborNet tree based on nuclear genomic sequencing. Strains were colored based on geographic isolation (North America, South America, Eurasia, Africa, Australia). **B)** WGS of strains were reference mapped to the 35 kbp ME49 apicoplast sequence (GenBank KE138841). 232 SNP positions were identified and a fasta file of the entire apicoplast sequence was constructed for each isolate. Apicoplast genomes were aligned to produce a maximum likelihood tree in the MEGA program with 1000 bootstrap support. Incongruence between the organellar and nuclear genomes grouped via the NeighborNet analysis support a model whereby recombination has occurred among the Type II isolates examined. **C)** Inset of NeighborNet tree in A focusing only on the Type II strains. Type II strains are significantly different from the other clades of strains in this analysis. Mitotic drift is apparent in the independent branching patterns of the Type II isolates. Meiotic recombination is shown in the reticulation between strains with interconnecting branches such as around ME49 and TgNgBr1.


A
0.01
COUG
MAS
TgCatBr5
TgCatBr9
RH-88
RH-JSR
GT1
CAST
VAND
VEG
CTG
3133
ARI
DAG
WTD1
TgSkunk
B41
3671
3675
TgGoatUs19
ME49
TgVaUs4
DEG
2987
TgNgBr1
CZ1
TgCatEg51
3131
TgGoatUs2
TgCatEg45
PRU
Goat_TAS
TgVvUs1
76K
TgDogEg11
I
II
III
X
B
GT1 apicoplast
RH-88 apicoplast
RH-JSR apicoplast
TgDogEg11 apicoplast
TgCatEg51 apicoplast
TgGoatUs2 apicoplast
Goat TAS apicoplast
2987 apicoplast
TgVvUs1 apicoplast
TgVaUs4 apicoplast
3131 apicoplast
TgNgBr1 apicoplast
TgGoatUs19 apicoplast
DEG apicoplast
CZ1 apicoplast
PRU apicoplast
76K apicoplast
TgCatEg45 apicoplast
3675 apicoplast
TgCatBr9 apicoplast
MAS apicoplast
TgCatBr5 apicoplast
WTD1 apicoplast
TgSkunk apicoplast
B41 apicoplast
3671 apicoplast
COUG apicoplast
ARI apicoplast
DAG apicoplast
ME49 apicoplast
3133 apicoplast
CAST apicoplast
CTG apicoplast
VEG apicoplast
VAND apicoplast
Africa 1
Africa 2
NA 1
NA 2
NA 3
NA 4
SA
Eurasia 1
Eurasia 2
NA 5
NA 6
0.001
C
0.001

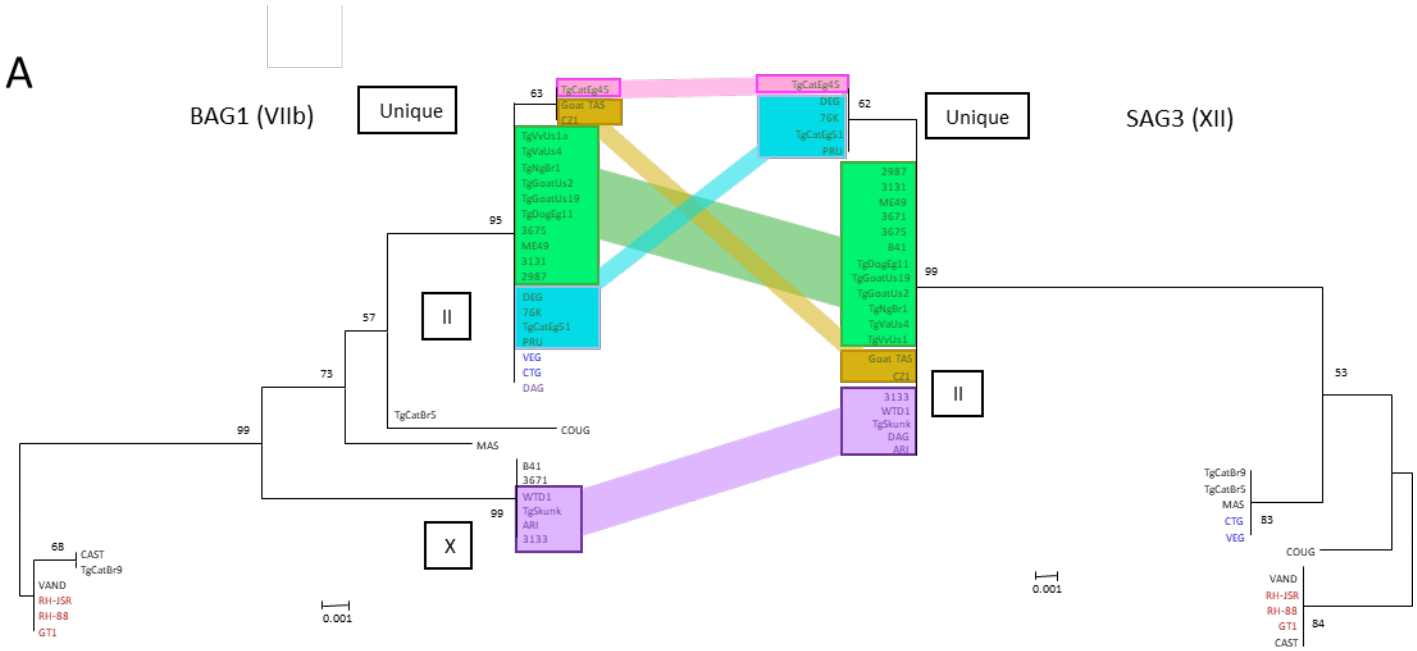
Marker Recombination is Indicative of Meiotic Recombination Across Type II Strains

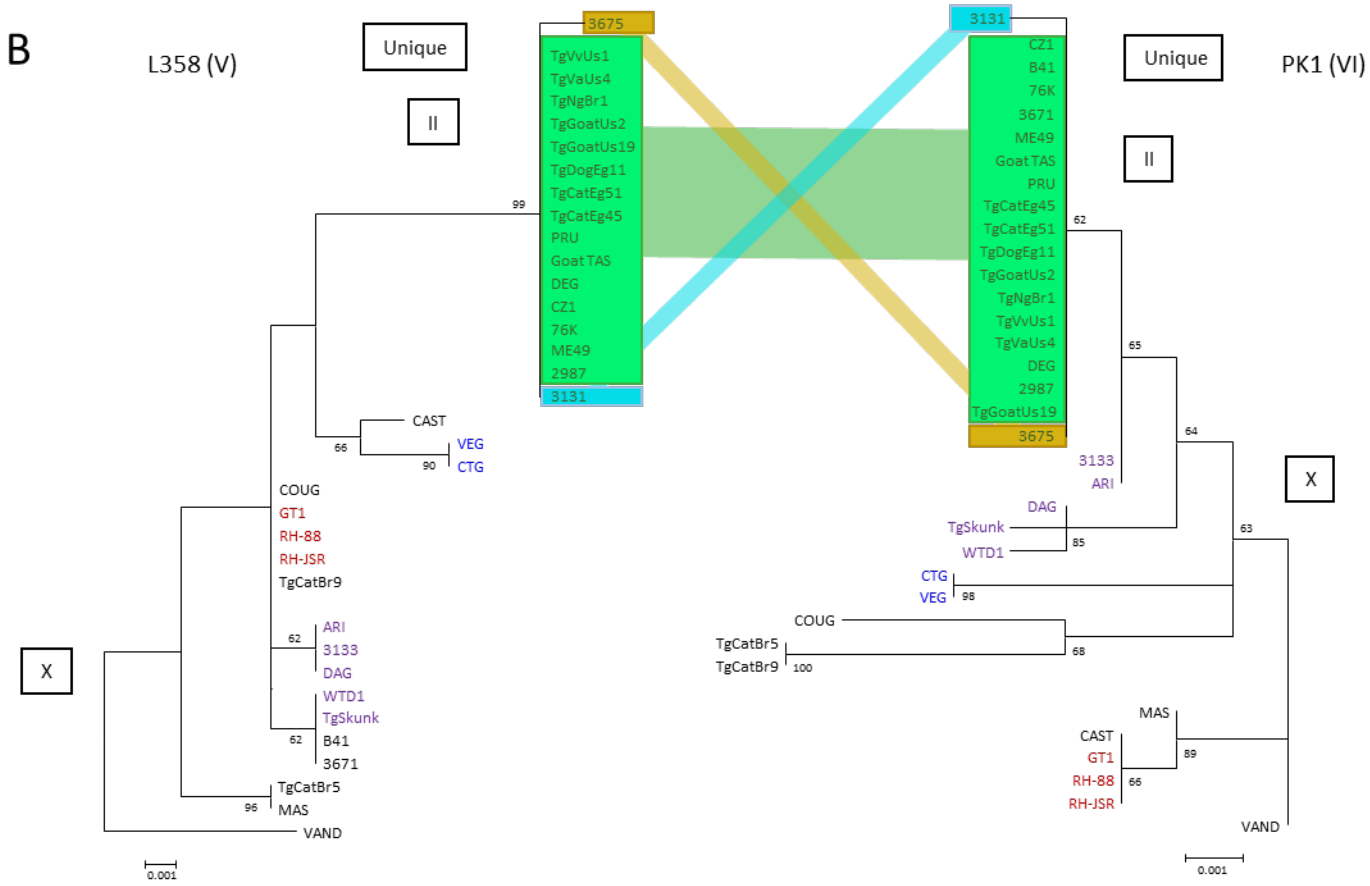
To investigate whether the genetic relationship between Type II strains is supported by a model of clonal expansion, sexual recombination, or some combination of both, individual maximum likelihood trees were created for each of the nuclear encoded single gene copy typing markers (Supplemental Table 1). To differentiate minor mitotic drift from independent evolution of a divergent allele, a bootstrap of 60% or greater was used to support a new allele designation. For each tree, all Type II strains were colored green while the I, III, and X lineages were colored red, blue, and purple respectively (Figure 3). Each locus bears a novel allele that has not been described in any of the canonical clonal strains or any of the 8 outgroup strains selected to assess genetic diversity within the species. Not only is this novel allele observed to possess greater than 60% bootstrap support and is divergent from the canonical Type II allele, but multiple strains bear the same allele at most loci. Importantly, no minor mutational drift was reported for each of the alleles resolved indicating a recent genetic origin. This data is parsimonious with sexual replication reshuffling a limited set of alleles within the Type II clade, presumably by unisexual mating. In total, 11 haplotypes were resolved across the 14 markers used to describe the 17 Type II strains (Supplemental Table 1A). Evidence for recombination is apparent at the majority of loci. For example, two alleles are resolved among Type II strains at both BAG1, a marker on chromosome VIIb, and SAG3, a marker on chromosome XII. The tree topology is congruent for 11 out of 17 strains, with ten versus one sharing either a canonical or novel Type II allele between the two markers, respectively. For the remaining 6 strains, crossing of the blue and orange lines depicts incongruence in the tree topologies and supports genetic hybridization as the most likely explanation (Figure 3A) for these two unlinked markers. Again, for two other unlinked markers versus L358, on chromosome V, and PK1, on chromosome VI, two alleles are

resolved, and two strains 3675 and 3131 are incongruent, visualized by the crossing of the orange and blue lines (Figure 3B). The crossing of the blue, orange, and green lines highlight the lack of congruence between the two sets of marker loci. Such incongruence across the Type II genomes can only be obtained via sexual recombination within the Type II clade of strains, arguing against the clonal theory of expansion for the Type II lineage. These recombination patterns were unexpected for a clonal lineage thought to be asexually expanding through the host population. To determine the extent to which recombination is shaping the genetic relationship among Type II strains, whole genome sequencing was pursued.

Figure 2-3: Segregation of unlinked and recombination among linked markers identified genetic hybridization within Type II strains

Comparison of sequence typing markers shown via maximum likelihood trees with 1000 bootstrap support at each marker. Strain identity here is characterized by the PCR-RFLP designation in Figure 1A. ME49 classifies the canonical Type II allele and WTD1 classifies the Type X allele. New alleles at each marker are differentiated based on 60% or greater bootstrap support. **A)** Comparison of unlinked markers: BAG1 on chromosome VIIb and SAG3 on chromosome XII. Green highlights strains that have Type II alleles at both loci. Purple highlights strains that differentiate Type X strains at BAG1. Pink highlights strains that contain a unique allele at both loci. The crossing of orange and blue highlight sexual recombination between the Type II strains as the alleles highlighted bear a unique allele at one locus and a canonical Type II allele at the other locus. **B)** Unlinked markers: L358 (chromosome V) compared to PK1 (chromosome VI). As above, green highlights strains that have Type II alleles at both loci. Orange and blue highlight strains that have undergone sexual recombination and possess a novel allele at one locus but a canonical Type II allele at the other locus.


A
BAG1 (VIIb)
Unique
II
X
SAG3 (XII)
Unique
II
0.001
0.001


B
L358 (V)
Unique
II
X
PK1 (VI)
Unique
II
X
0.001
0.001

The Type II Population Clades into Several Distinct Sub-groups by Genomic Inheritance

In order to interrogate the distribution of polymorphism and whether the novel alleles identified in the phylogenetic trees (that have genetic ancestry closest to Type II) exist in discrete haploblocks, or are randomly distributed across the genome, pairwise SNP diversity plots were generated for all Type II isolates, as well as reference Type I and III strains that are known to have previously recombined with Type II (Boyle et al. 2006). Although SNP diversity fingerprinting does not report the SNP identity, density patterns display strain differences across the genome as a whole, allowing visualization of a fingerprint of the diversity encoded within the genome all at once. Strains were reference mapped to ME49 and differences were shown as a quantification of diversity in 100 kbp windows across the genomes. The Type I RH reference strain possessed a Type II chromosome Ia and IV, and a Type II admixture block at the left side of chromosome VIIa, and right end of chromosome XI. The Type III VEG strain displayed a similar admixture with Type II sequence across roughly 40% of is genome, as expected based on previous admixture characterizations (Figure 4A). A select number of Type II strains showed similar admixture diversity as Type I and III, although the SNP density was approximately one log lower (3-5 SNPs per 10 kb block) in the regions that were different from ME49 reference, consistent with the diversity observed at the individual marker loci that were Sanger sequenced. To determine whether this diversity was in large haploblocks synonymous with recombination, and whether this was a common finding across the entire Type II clade, SNP density fingerprints were created for all Type II strains. The plots were grouped by geographic isolation (Figure 4B). Strong correlations between geographic isolation and genomic diversity were apparent indicating that Type II strains from different geographic locations can be differentiated from each other by

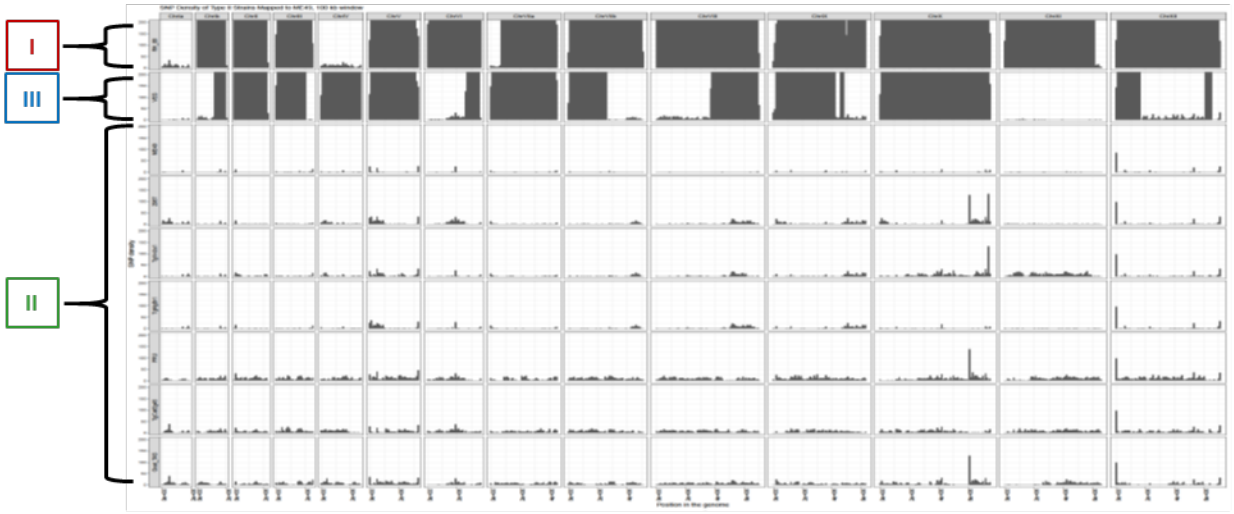
the accumulation of shared SNPs that are geographically restricted. Importantly, all of the North American isolates appear to be genome-wide admixtures that share distinct portions of their genomes with ME49, and other regions that share closest ancestry with Eurasian, African and Australian isolates. The picture that emerges is that the majority of Type II strains are distinct genetic admixtures of the SNP diversity present within the Type II clade. This is also captured by the inheritance of the apicoplast genome. For example, the Australian strain GoatTAS is distinct genome-wide, except for a large haploblock of sequence in the first third of chromosome VI which shares ancestry with North American strains. Not surprisingly, the Goat TAS apicoplast genome clades with the North America 2 apicoplast, indicating that this line has crossed with a Type II strain from North America (Figure 2B). Conversely, while Eurasian isolates CZ1 and PRU group closely by apicoplast phylogeny, their SNP densities are distinct and CZ1 chromosome V shares ancestry with North American strains. SNP density fingerprints within the North American strains correlated well with the diversity seen in the apicoplast tree but showed more SNP diversity haploblocks than any other geographical sub-group, indicating that multiple sub-groups exist within this group of geographically-related Type II strains. Importantly, the SNP diversity among the North American Type II strains was inherited in discrete haploblocks rather than being randomly distributed across the genome, which strongly supported an admixture model for expansion of the Type II lineage by unisexual mating.

While the pairwise SNP diversity plots can identify haploblocks that are different from reference, these SNP plots do not identify the ancestry of non-reference haploblocks. We next utilized PopNet to identify the number of discrete ancestries comprised within the Type II clade, and to map these ancestries across each genome to identify strains that are most similar to each other (Figure 4C). PopNet identifies chromosomal regions of shared genetic inheritance by

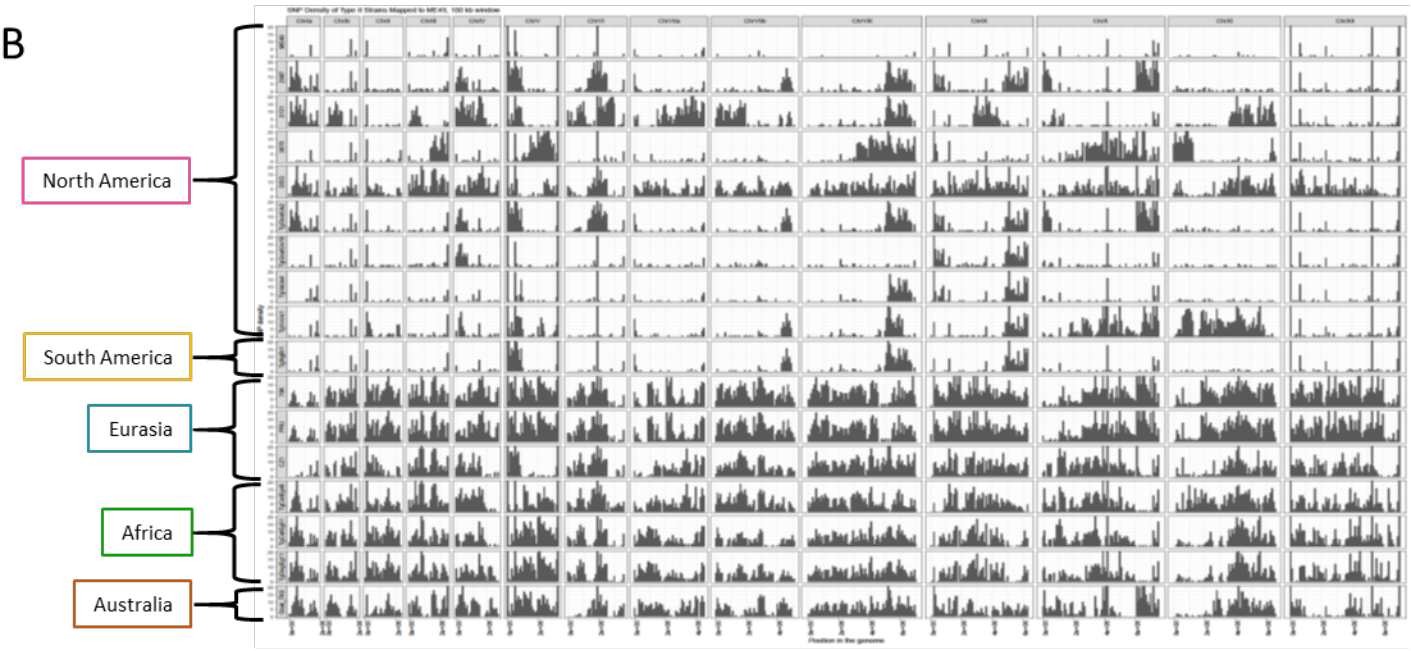
looking at recombination blocks based on clustering shared polymorphisms across all of the input genomes. Chromosomes are painted based on blocks of shared sequence. Chromosome painting of these Type II strains readily visualized recombined sequence haploblocks of distinct ancestry. For instance, the North American strains, TgGoatUs2 and 2987, had similar SNP density fingerprints (Figure 4B), and both share a large haploblock on chromosome Ia with Type I strains that is not present in the other Type II strains, indicating that these two strains are of mixed ancestry. Type II strains that share the most ancestry are indicated by the bolder connecting lines. Interestingly, one North American isolate, TgVvUs1, appears to have haploblocks of sequence in common with multiple lineages of Type II from diverse global populations. This mixed inheritance within a Type II strains strongly supports recombination occurring within the Type II clade as the inheritance of discrete haploblocks of shared ancestry could only be generated by unisexual mating between related genomes.

Figure 2-4: Characterization of genomes of Type II strains via SNP density and PopNet shows diversity is inherent across the Type II strains and indicates diverse haploblocks are recombining across the Type II genomes.
A) Pairwise SNP diversity plots of the Type I and III clonal lineages as well as a number of representative strains of Type II. These plots were created from WGS SNPs mapped to ME49. Each row displays the genome of one strain of Type II. Columns separate the chromosomes based on chromosome length. Each vertical bar represents the number of SNPs in 100 kbp window of sequence with the tallest bars representing areas in this strain which contain the highest quantity of SNPs that differ from the Type II ME49 reference strain sequencing, representing greater diversity in that area of the genome. **B)** SNP density plots for all Type II strains grouped by geographical isolation of the strains. Y-axes are constrained to 0-20 SNPs per window to allow better visualization of Type II strains. **C)** PopNet analysis of the WGS of the Type II and X strains with Type I, III, and X strains for reference. All chromosomes of these strains are concatenated into a single circular representation of the genome. Chromosomes are painted based on their shared ancestry within groups of strains. These clusters are shown via the color of the circle behind the strain which indicates the clustering of the majority of its genome (Type II-green, Type X-purple, Type I-cyan, Type III-blue). Lines between strains indicate interrelatedness between these particular strains with bolder lines indicating a stronger likelihood of clustering together as with the PRU and 76K strains.

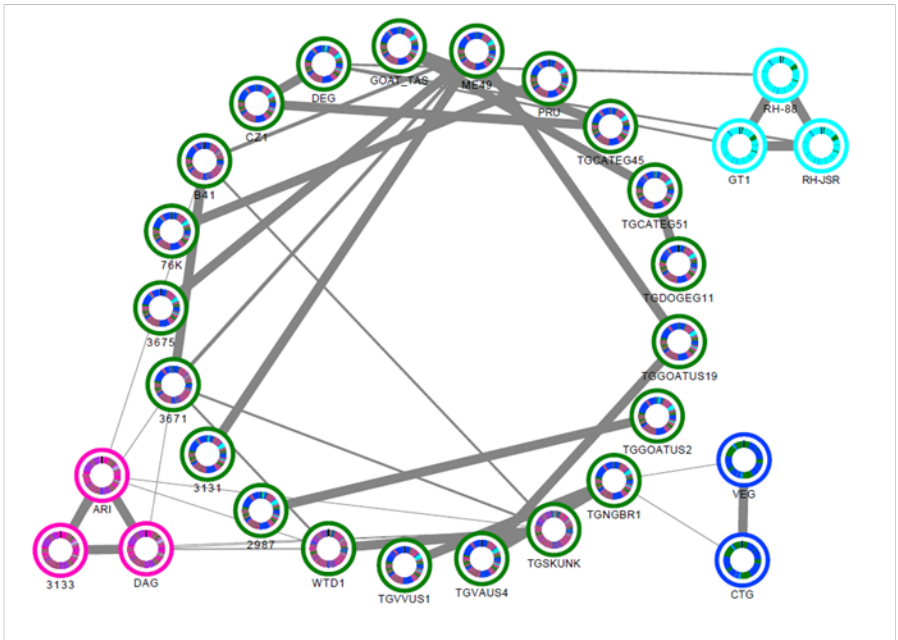
A


I
III
II

B


North America
South America
Eurasia
Africa
Australia

C

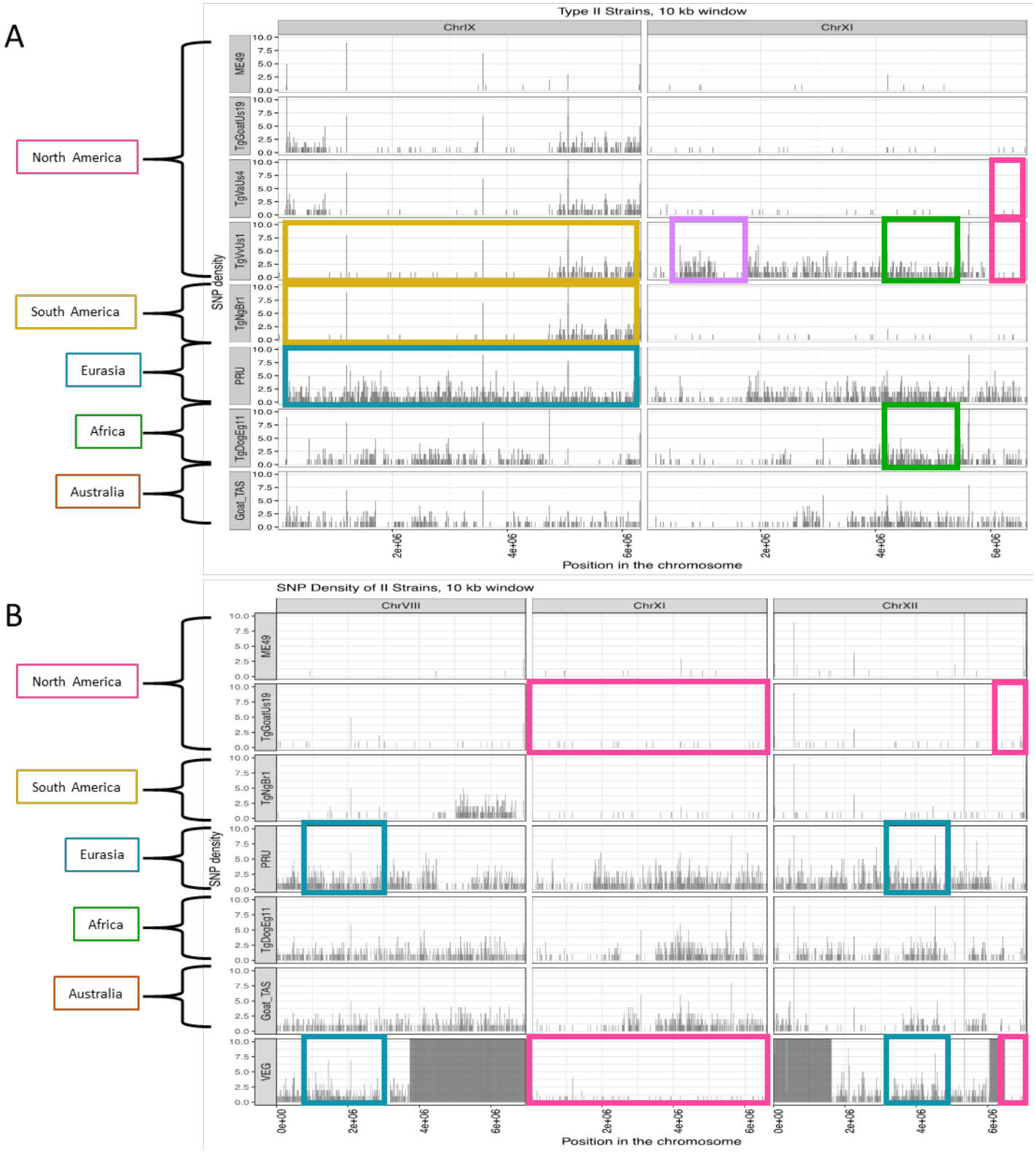

DEG
GOAT_TAS
ME49
PRU
RH-88
C21
TGCATEG45
GT1
RH-JSR
B41
TGCATEG51
76K
TGDOGEG11
3675
TGGOATUS19
3671
TGGOATUS2
3131
VEG
ARI
TGNGBR1
2987
TGSKUNK
3133
DAG
WTD1
TGVVUS1
TGVAUS4
CTG

SNP Fingerprinting Displays Type II Diversity and Recombination Between Strains

After observing genetic admixture of distinct haploblocks from geographically isolated strains within a single North American Type II genome, further investigation of the TgVvUs1 haploblocks was necessary to understand the extent of chromosomal recombination across this strain. A subset of 7 representative isolates from across the geographic range of sequenced strains were interrogated to characterize the origin of these discrete haploblocks. The SNP fingerprint of specific TgVvUs1 haploblocks were compared with fingerprints of other analyzed strains to show that the TgVvUs1 SNP density pattern across the length of chromosome IX was shared with the South American strain TgNgBr1. This pattern of inheritance was not observed in other North American isolates, indicating that this haploblock is not common in North America but has likely admixed into the TgVvUs1 genome via sexual recombination between North and South American Type II strains. The SNPs in this chromosome were extracted from the WGS of all strains previously characterized in this study and were analyzed in a maximum likelihood tree (Supplemental Figure 2A). Both strains clustered on the same 100 bootstrapped tree branch, which supported the finding that chromosome IX of TgVvUs1 was genetically most similar to chromosome IX of the South American isolate. As expected for a North American strain, the last 1,400,000 bp of chromosome XI shared ancestry with other North American strains (Supplemental Figure 2D). However, the middle portion of the chromosome from 4.2-5.4 Mbp displayed inheritance consistent with African isolates such as TgDogEg11 (Supplemental Figure 2C). TgVvUs1 also displayed some unique evolution at the proximal end of Chromosome XI where the sequence grouped separately from all other Type II strains, both in SNP density and phylogenetic analysis (Supplemental Figure 2B), which may be evidence of ongoing mitotic drift or introgression of another ancestry.

The Type III *Toxoplasma* strain, VEG, is known to be the product of a sexual cross between a Type II strain and another distinct ancestry previously referred to as β (Boyle et al. 2006). To determine the ancestry of the type II haploblocks present in VEG, a SNP diversity plot was generated, and the SNP fingerprinting pattern was similar to that observed for TgVvUs1 (Figure 5B). The proximal 3.8 Mbp of chromosome VIII possessed a Type II haploblock and the maximum likelihood tree generated in this region claded with Eurasian Type II strains (Supplemental Figure 3A/B). Conversely, SNP fingerprint analysis and maximum likelihood trees supported the entirety of chromosome XI as most similar to North American Type II strains (Figure 5B and Supplemental Figure 3C). However, the VEG chromosome XII had a mixed heritage (Figure 5B). In the 3.6-4.6 Mbp region it was most closely related to the Eurasian strains (Supplemental Figure 3D) whereas the distal 6.5-7.1 Mbp of the chromosome grouped more tightly with North American Type II strains (Figure 5B and Supplemental Figure 3E) suggesting that the β parent likely crossed multiple times with different Type II lines, or once with a recombinant that was similar to TgVvUs1 to create that Type III VEG strain.

Figure 2-5: Recombination observed across haploblocks of SNP density identified in both Type II and Type III strains.
A subset of the SNP density fingerprints shown in Figure 4A are highlighted here to display recombination that has occurred across the genomes of the Type II TgVvUs1 strain, isolated from a North American fox, and the Type III VEG strain. SNP density plots are shown using a sliding 10 kbp window to allow for better resolution of discrete haploblocks of SNP density on the selected chromosomes. Y-axes are restricted to 0-10 SNPs to resolve the low frequency of divergence within the Type II clade. Representatives strains of each geographical isolation are included for reference to the isolate being interrogated. Strains were colored as previously based on geographical isolation (North America, South America, Eurasia, Africa, Australia). **A)** North American strain, TgVvUs1 SNP density fingerprint across chromosomes IX and XI compared to geographically representative strains. Chromosome IX highlighted in yellow mirrors the SNP density of South American strain TgNgBr1. Chromosome XI highlights the mixed inheritance pattern in TgVvUs1. The last 1.4 Mbp of chromosome XI (highlighted in pink) has a shared ancestry with North American strains, whereas the region from 4.2-5.4 Mbp (highlighted in green) has a shared ancestry with African strains *(e.g.,* TgDogEg11). Purple highlights the haploblock region from 1.2-1.8 Mbp where TgVvUs1 appears to have a distinct ancestry from all other Type II strains assayed. Chromosomal trees supporting these inheritances are shown in Supplementary Figure 2. **B)** VEG SNP density fingerprint across chromosomes VIII, XI, and XII compared to geographically representative strains. The first 3.8 Mbp of chromosome VIII (highlighted in blue) has a SNP density fingerprint similar to the Eurasian DEG strain whereas the entirety of chromosome XI (highlighted in pink) shares ancestry with North American Type II strains such as TgGoatUs19. Chromosome XII identifies recombination between different ancestries. Specifically, the SNP density from 3.6-4.6 Mbp (blue) encodes a divergent Type II ancestry that is most closely related to the Eurasian PRU strain, whereas the region from 6.5-7.1 Mbp (pink) shares ancestry with North American Type II strains (*e.g.,* TgGoatUs19). Chromosomal trees supporting these inheritances are shown in Supplementary Figure 3.


A
North America
South America
Eurasia
Africa
Australia
Type II Strains, 10 kb window
ChrIX
ChrXI
ME49
TgGoatUs19
TgVatUs4
TgVvUs1
TgNgBr1
PRU
TgDogEg11
Goat_TAS
SNP density
Position in the chromosome
B
SNP Density of II Strains, 10 kb window
ChrVIII
ChrXI
ChrXII
VEG

Discussion

Toxoplasma and the Clonal Theory of Expansion

Sexual recombination is known to alter biological potential within a parasitic population, allowing selection for advantageous adaptations that are maintained cryptically and when sexually recombined can confer on the progeny new combinations of virulence genes that enhance transmissibility and extend a parasites host range. This adaptability is especially important for generalist parasites, like *Toxoplasma*. Provided the opportunity to sexually mate, *Toxoplasma* can produce upwards of 800 million highly-infectious, genetically recombinant progeny, all of which are capable of productively infecting any warm-blooded animal (Torrey and Yolken 2013). This fecundity would suggest that the sexual cycle of *Toxoplasma* should aide in the ability of this parasite to expand across a broad host range. However, previous work has suggested that the *Toxoplasma* population is expanding primarily asexually, which would limit punctuated adaptations that could be rapidly recombined within the population by its sexual cycle. The apparent clonality of the *Toxoplasma* population led to the clonal theory of propagation; a theory proposing that *Toxoplasma* strains propagate in nature exclusively by asexual expansion, and that sexual recombination was sufficiently rare to not impact the clonal population structure (Sibley and Ajioka 2008; Tibayrenc and Ayala 2002, 2014). The clonal theory of propagation states that mitotic drift is occurring due to polymerase error during replication, but that sexual recombination is not sufficient to break up the clonal population structure. The theory was based largely on a limited set of genetic markers that were in linkage disequilibrium and suggested the existence of widespread identical genotypes. In the age of genome sequencing, it is possible to revisit the theory of clonal propagation to test the

sufficiency of the model, when virtually an entire genome of markers are interrogated. This thesis assayed at full genome resolution a collection of strains from the globally distributed Type II clonal lineage to determine the extent to which isolates within this lineage, are expanding asexually or sexually. The Type II clade was selected to address the possibility that *Toxoplasma's* clonal lineages may possess greater diversity than previously envisaged, using higher resolution markers and whole genome sequencing to determine the contribution of the sexual and asexual lifecycles in shaping the clonal population structure of *Toxoplasma.*

Diversity Inherent in Typing Markers of the Type II Clonal Lineage

Work shown here tested the sufficiency of the current set of typing markers to accurately predict strains that belong to a clonal clade. The Type II clade is defined by a restriction digest pattern at ten ToxoDB PCR-RFLP characterization markers. We selected 17 strains that were known to be in linkage disequilibrium and genotyped as Type II strains by PCR-RFLP analyses. The first test was to expand the number of markers and DNA sequence both the PCR-RFLP markers as well as 4 additional typing markers that were selected because they were under either neutral, homogenizing and diversifying selection. Our results indicated that within the Type II clade, multiple, genetically distinct Type II sub-groups were present across the globe once analyzed by whole genome DNA sequencing. This genetic diversity may begin to explain the differences observed in clinical disease within the Type II clade.

The diversity within the Type II clade was not entirely unexpected as this clade is known to be comprised of two distinct PCR-RFLP genotypes grouped into one clade of strains. The one known SNP difference is in the maternally-inherited apicoplast organellar genome marker (APICO) which is not nuclear-encoded, and hence was not considered in the Type II lineage

designation (Su, Zhang, and Dubey 2006). In this study, DNA sequencing of the full 35kb apicoplast genome on which APICO is located identified considerable genetic diversity within the maternally inherited, organellar genome. Genomic diversity using the expanded set of sequenced nuclear encoded markers within our cohort of 17 Type II strains suggested greater diversity among Type II strains (Ajzenberg et al. 2009; Verma et al. 2015). Based on marker sequencing, five distinct haplotypes were identified within the Type II strains examined and these preliminary results indicated that whole genome sequencing would be needed to determine the true extent of diversity within the Type II clade.

Incongruence Between Nuclear and Organellar Genomes Indicates Sexual Recombination Occurs Within the Type II Clade

Diversity within these strains could be from polymerase error during mitotic replication (genetic drift) or from recombination, which admixes sequence haploblocks bearing distinct ancestries and can cause incongruence between the apparent ancestries of nuclear-meiotically dividing and organellar-mitotically dividing genomes. To investigate the degree to which isolates within the Type II clade were expanding asexually versus sexually, a phylogenetic analysis was carried out to determine whether incongruence among tree topologies existed between the nuclear and organellar genomes of Type II. The apicoplast is thought to evolve exclusively by mitotic drift and does not participate in meiotic sexual recombination, hence, an incongruence between the nuclear versus the organellar genome would indicate that these genomes have sexually recombined. The network analysis for the genomes of these strains (Figure 2A) displays known clonal lineages on individual branches with individual isolates of these lineages clustered closely together at the end of each branch, as expected for lineages with a small degree of mitotic

variation (Lorenzi et al. 2016). Type II strains all claded together on a single branch, supporting their close genetic relationship. However, despite the apparent clonality of the Type II strains (Figure 2A), the mapping of these strains onto the apicoplast genome identified 10 haplotypes that resolved into six well-supported (with >60% bootstrap support) apicoplast genomes. Interestingly, the six apicoplast genomes did not all clade with the same five groups that were defined by the sequencing the nuclear encoded typing markers. The only way that this incongruence could occur is by genetic hybridization which admixes nuclear genomes but does not affect the maternal inheritance of the apicoplast genome which undergoes mutation by genetic drift (Ferguson et al. 2005).

To determine whether the Type II strains are undergoing unisexual expansion (intra-clade genetic recombination) and reshuffling diversity within the Type II clade, as suggested by the typing markers, WGS was performed on 17 strains Type II strains to better resolve the allelic diversity discovered and determine whether this diversity is distributed in large haploblocks synonymous with genetic hybridization. In fact, the majority of the typing markers possessed at least one novel allele that closely resembled the ME49 canonical Type II allele but was considered sufficiently different to be evolving independently based on bootstrap support greater than 60%, a cut-off that was previously used to differentiate the Type II and X clades. Furthermore, the novel II-related alleles were limited and did not possess strain-specific SNPs as would be expected if these alleles were ancestral and accumulating mutations by polymerase error or mitotic drift.

Type II Strains Display Geographical Inheritance and Sexual Recombination Across the Genome

SNP density plots established that the Type II strains resolved into several discrete subgroups based on their geographic origin. However, both the Eurasian and North American groups had at least one strain that contained SNP diversity fingerprints not seen in the rest of the geographic group. North American strains were the most widely sampled and also the most diverse geographic group within the analysis. Correspondingly, the North American Type II isolates (such as 3131 and TgGoatUs2) displayed more in-group SNP fingerprint diversity than the other geographic groups, indicating that there existed multiple subgroups of Type II within North America. Prevalent haploblock recombination across these North American sub-groups indicated that sexual recombination as well as mitotic drift was shaping the Type II population. Indeed, the geographical subgroups observed by the SNP diversity analyses were supported by the PopNet chromosome painting analysis.

The North American strain TgVvUs1 was a good example of full genome admixture between different clades of Type II. TgVvUs1 was shown to share sequence haploblocks that were either of South American, African, or North American descent, indicating that this isolate has undergone several unisexual genetic hybridizations with a diverse array of geographically isolated groups within the Type II clade. Importantly, the majority of the genome's SNP fingerprint clades with the North American Type II sub-group, and this strain has a limited number of recombination points that are readily observed across the genome, which is not parsimonious with this strain evolving from a common ancestor but is rather a product of multiple genetic hybridization events between divergent strains of Type II. Although only 17 strains were sampled across the global Type II population, and at least four of these (CZ1,

TgVvUs1, 3131, TgCatEg51) were found via SNP density fingerprinting and phylogenetic verification to be recombined progeny from unisexual crosses between Type II strains, indicating that sexual recombination is likely more prevalent than has previously been characterized, especially since no previous study has interrogated any clonal lineage at whole genome resolution (Sibley et al. 2009; Grigg and Sundar 2009). Recombination within a clade masks the frequency of recombination within the *Toxoplasma* population, especially because the frequency of sexual crossing is generally assayed by a limited set of PCR-RFLP markers that were designed to distinguish the original three clonal lineages rather than distinguishing intra-lineage diversity (Shwab et al. 2014; Khan, Dubey, et al. 2011; Boyle et al. 2006; Minot et al. 2012). It is likely that further interrogation of the Type II clade will yield additional diversity and evidence of recombination across this clade of strains, especially with the addition of more South American isolates where sexual recombination is known to be an integral part of population transmission dynamics (Khan, Dubey, et al. 2011).

The diversity observed within the Type II clade is recapitulated within the Type II regions of the Type III strain VEG. Type III strains are progeny of an ancient cross between an unknown β ancestor and a Type II ancestral strain (Boyle et al. 2006). SNP density fingerprinting of the Type II haploblocks within the Type III VEG strain showed inheritance of both North American and Eurasian ancestry. Based on the limited diversity and strong segregation of the Type II haploblocks, the mixed inheritance of the Type II ancestry within the Type III clade was likely derived from sexual crossing between a mixed North American/Eurasian Type II strain and an unknown β strain rather than subsequent interbreeding with multiple Type II strains. The mixed inheritance of the Type II strain agrees with previous studies where Type III strains contained two distinct sub-groups of the Type II lineage based on

sequence comparisons of haploblocks across the population (Minot et al. 2012). Recombination within the Type III clade as well as the Type II clade supports a model whereby genetic hybridization is shaping the population structure of *Toxoplasma,* at least since VEG was derived ~5000 years ago (Boyle et al. 2006). Additionally, this ancient recombination indicates that sexual recombination between the Type II isolates is not a recent adaptation of this clade. Whether this diversity of genomic inheritance would be recapitulated in other clonal lineages would inform the mechanisms by which *Toxoplasma* maintains its apparently clonal population structure.

Sexual Recombination is Frequent yet Biased Within Type II Strains, and Likely the *Toxoplasma* Population

Work done herein indicates that the clonal expansion theory of *Toxoplasma* evolution is insufficient to characterize the population dynamics of this generalist pathogen. The Type II clade of strains is not solely expanding asexually, but rather replicates via a combination of asexual propagation and unisexual mating (between closely related strains) which work in tandem to maintain a population structure that appears clonal when surveyed by limited resolution markers. The clonal expansion model misrepresents the dynamic adaptability of the *Toxoplasma* population as a whole. In fact, unisexual expansion is more frequently utilized by *Toxoplasma* than previously envisaged for the Type II clade. The lack of outcrossing observed within the Type II strains characterized here may indicate a preference toward similar sexual partner mating. This preference needs to be further interrogated as the specifics of sexual recombination in the feline host are still unclear. With *Toxoplasma*, although sexual recombination is optional for replication, it increases biological potential and infective capacity

and allows for the diversification of the population across the host range even when avirulent parents are recombined (Grigg, Bonnefoy, et al. 2001). An inaccurate estimation of the frequency of sexual recombination across the *Toxoplasma* population skews the expected population dynamics and alters estimates of times of divergence from a common ancestor significantly. This is especially true for *Toxoplasma,* where the calculated *Plasmodium* mutation rate, which has an obligate sexual cycle of replication, is used to estimate divergence from a common ancestor (Boyle et al. 2006; Fux et al. 2007; Shwab et al. 2014). Low marker resolution is the most likely cause of this underestimation of diversity within the *Toxoplasma* population. Using low-resolution genetic typing, unisexual expansion between closely-related strains, such as between Type II isolates, has been erroneously mistaken for asexual expansion in the absence of other genomic hallmarks of recombination. This misclassification of how *Toxoplasma* is expanding in nature underestimates the contribution of felines in the expansion, evolution and transmission potential of this highly successful protozoan pathogen. To determine the respective contributions of genetic hybridization by unisexual or uniparental mating within the population, an investigation of an isolated population of strains needs to be done over a time course with whole genome sequencing performed on all strains isolated to determine the contribution of each mechanism to the population dynamics and transmission of this highly prevalent, generalist parasite.

Materials and Methods

Toxoplasma Strain Collection Typing

Seventeen *Toxoplasma gondii* strains initially characterized as part of the Type II lineage were collected from unique geographical locations across the globe (Supplemental Figure 1). 3 Type I, 2 Type III, 5 Type X, and 8 outgroup strains were also included in this collection (Figure 1A) (Lorenzi et al. 2016). All strains had previously been typed by the laboratory in which they were isolated using a the standard ToxoDB Su genotyping PCR-RFLP typing markers (Su, Zhang, and Dubey 2006; Su et al. 2012). Strains were grown on human foreskin fibroblast (HFF) monolayers fed on Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 2mM glutamine, and treated with gentamicin and penicillin/streptomycin. Tachyzoites were harvested at time of HFF monolayer lysis and filtered through a 3.0-micron polycarbonate filter to remove host cell contamination (Khan, Dubey, et al. 2011). DNA from the filtered parasites was extracted using Qiagen's Blood and Tissue DNeasy extraction kit. DNA was used to genotype strains using the standard ten PCR-RFLP markers (SAG1, SAG2, SAG3, BTUB, GRA6, L358, PK1, c22-8, c29-2 and APICO) to characterize new strains into the ToxoDB PCR-RFLP groups (Figure 1A) (Shwab et al. 2014; Su, Zhang, and Dubey 2006; Sundar et al. 2008). Further classification was done using additional PCR markers common to the Grigg laboratory (UPRT, EF1, BAG1, and SRS2) (Supplemental Table 1) (Fux et al. 2007; Parameswaran et al. 2010). Each marker was first amplified by PCR using Taqman Ampli-Taq polymerase and an Eppendorf thermocycler as previously described (Parameswaran et al. 2010; Su, Zhang, and Dubey 2006). The PCR products were then run on a 1% agarose gel to verify amplification of the correct product. All PCR products were restriction enzyme digested as

described previously (Su, Zhang, and Dubey 2006; Grigg and Sundar 2009; Parameswaran et al. 2010). The PCR products were Sanger sequenced and sequences analyzed using DNASTAR's Lasergene software version 14.

Whole Genome Sequencing of Type II Strains

DNA collected from *Toxoplasma* strains was sent to Rocky Mountain Laboratories and whole genome sequenced on a HiSeq machine using Illumina technology. All Type II strains were reference mapped to the ToxoDB Version 8.2 ME49 assembled genome or the ME49 assembled apicoplast genome KE138841.1. This was done according to best practices (as recommended by BWA and GATK) using BWA 0.7.5a to align the raw reads and GATK 3.7 with Picard 1.131 to quality control and filter the reads into a genome alignment (Van der Auwera et al. 2013). Strains whose average coverage across the genome was less than 5X using the VCFtools 0.1.8a genome coverage command were not included in subsequent analyses (Danecek et al. 2011). This coverage density was necessary to investigate sequence diversity across closely related strains, such as within the Type II clade. SNPs from whole genome sequenced strains were identified using the gVCF combination method in GATK (Van der Auwera et al. 2013). Briefly, a stand_call_conf of 30.0 and nct of 10 in GATK was used to call only haploid SNPs across the genomes.

Phylogenetic Tree Creation from Markers and Genomes

Sequences of the typing markers and the apicoplast genomes were aligned using Clustal X 2.1 with default settings (Larkin et al. 2007). For each marker, the alignment was input into MEGA 7 to create a maximum likelihood tree using Tamura-Nei model distance analysis with uniform rates of substitution across all sites, and 1000 bootstrap support for all branch points (Tamura et al. 2013). Bootstrap support equal to or greater than 60% distinguished novel alleles from mitotically drifted alleles because this level of bootstrap differentiation differentiates the closely related Type X and Type II lineages on single marker phylogenetic trees (Khan, Dubey, et al. 2011). All typing marker trees were rooted on the most genetically diverse strain included in this analysis, VAND. Phylogenetic trees of only the Type II and X strains were rooted on COUG, a known ancestor to the Type II clade (Minot et al. 2012).

Maximum likelihood trees for whole and partial chromosomes were created in much the same manner as those created for typing markers. Genetic variants determined from reference mapping were parsed according to their genomic location and a custom script was used to create fasta sequences containing only locations where a polymorphism was present in one or more strains. Due to the manner of their generation from an aligned genome sequence, these fasta files did not require alignment and were input directly into MEGA 7 to create maximum likelihood trees. Tree parameters are as described above, but only 100 bootstrap replicates were used because the extended sequence lengths and previous genomic alignment ensured the accuracy of these comparisons without the additional bootstrap support given to shorter marker alignments.

To create a NeighborNet tree based solely on the genomic SNPs, genomic polymorphism fasta files derived above were input into SplitsTree4 and default program parameters were used with 1000 bootstrap support for branch length (Huson and Bryant 2006).

SNP Density Fingerprint Analysis of Closely Related Strains of Toxoplasma

Using GATK and VCFtools, a VCF file containing SNP covering the whole genome for all whole genome sequenced strains was converted to tabular format containing only biallelic SNPs with no long insertions or deletions (Van der Auwera et al. 2013; Danecek et al. 2011). A series of custom R scripts were generated based off of the scripts used to map the location of the sinefungin resistance gene (Behnke, Khan, and Sibley 2015). If any, one strain in the VCF analysis has a polymorphism at a set locus, all strains nucleotide sequence is recorded at that locus, regardless of if they are identical to the reference strain at that location. These custom scripts extracted SNP positions and identities where the strain in question differs from the reference sequence, such that the data from each strain only contains positions where the strain is non-reference. The number of SNPs within each strain of *Toxoplasma* was counted in a specified window, of 10 or 100 kbp length. The number of SNPs across a strain over a set genomic distance was given a limited y-axis (0-10 bp for 10 kbp windows, 0-20 bp for 100 kbp windows) to better display the SNP diversity inherent in Type II lineages. A larger scale as for more divergent strains such as VEG was not used because it would have eclipsed the number of Type II SNPs that differ within a window. Genomic region diversity or similarity was grouped based on geographic isolation of the strains and then on visual similarity of the SNP density plots which were confirmed by phylogenetic analysis of the region.

PopNet Analysis of Shared Inheritance Across the Type II Strains

All 17 Type II, 5 Type X (ARI, DAG, WTD, TgSkunk, and 3133), 3 Type I (RH-88, RH-JSR, and GT1), and 2 Type III (VEG and CTG) strains were selected for comparison and were included from the WGS VCF created from all the strains (detailed above) for analysis by PopNet. The SNP VCF of these strains was input into the PopNet program and run on previously determined parameters for *Toxoplasma* strains to determine their interrelatedness. This analysis was visualized using Cytoscape as previously detailed (Zhang et al. 2017).

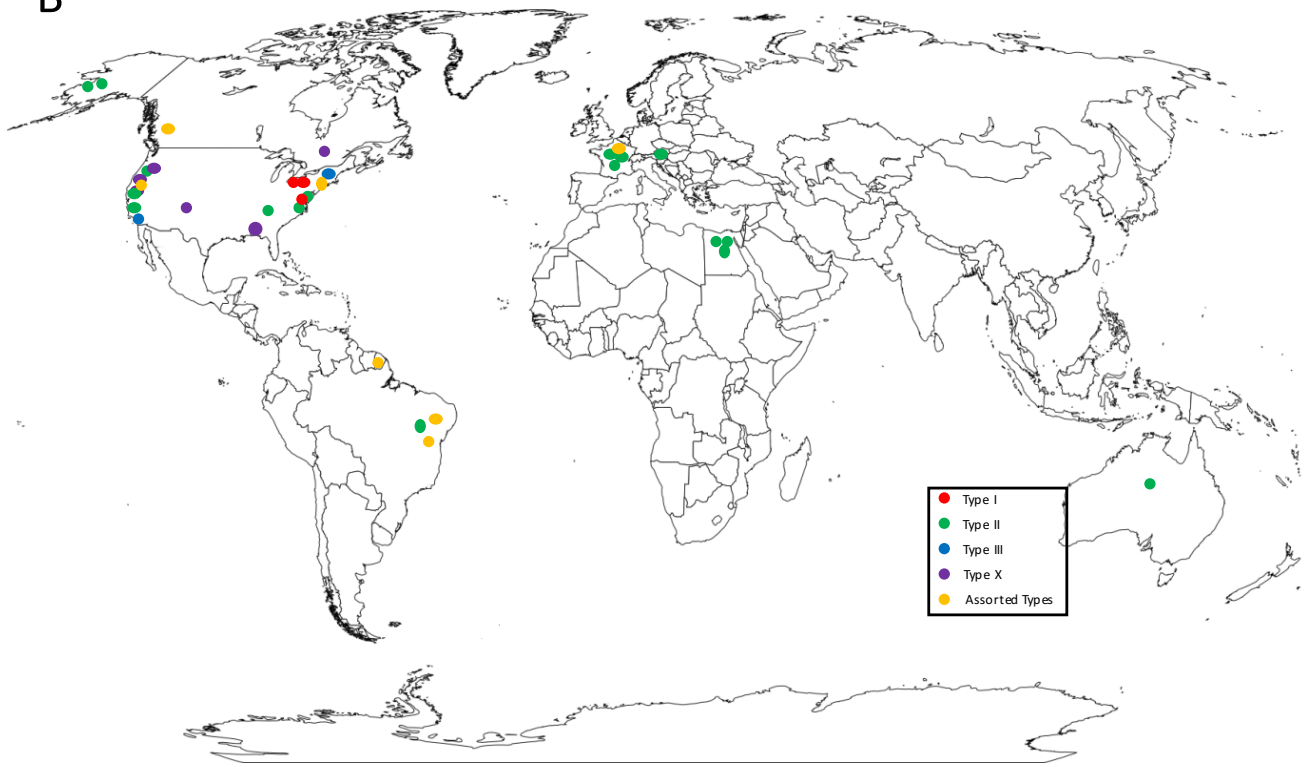
Supplemental Figures

Supplemental Figure 2-1: Strain isolation information
A) Strains used here are listed with PCR-RFLP typing designation, isolation host and location, and their sequencing identification number. **B)** Global isolation location is shown for all strains utilized here. Strains are represented by their PCR-RFLP typing designation as shown by dots in the legend with Type I-red, Type II-green, Type III-blue, Type X-purple, Outgroup strains-yellow.

A

Strain ID	Group by RFLP	Region	Sub-Region	Host	RML ID
GT1	I	USA	MD	Goat	SRR516419
RH-88	I	USA	OH	Human	SRR521957
RH-JSR	I	USA	OH	Human	SRR521559
Goat TAS	II	Australia		Goat	8206
TgNgBr1	II	Brazil			6591
CZ1	II	Czech Republic		Human	3893
TgCatEg45	II	Egypt		Cat	6544
TgCatEg51	II	Egypt		Cat	6588
TgDogEg11	II	Egypt		Dog	6590
76K	II	France	Tours		4818
DEG	II	France	Limoges	Human	7595
PRU	II	France	Limoges	Human	SRR350739
2987	II	USA	CA	Sea Otter	4808
3131	II	USA	CA	Sea Otter	4809
DAG	II	USA		Human	8476
TgGoatUs19	II	USA	MD	Goat	4824
TgGoatUs2	II	USA	MD	Goat	4820
TgVaUs4	II	USA	AK	Fox	4828
TgVvUs1	II	USA	AK	Fox	4829
3142XB7_1	II (ME49)	USA	CA	Sheep	6135
VEG	III	USA	CA	Human	SRR516406
CTG	III	USA	NH	Cat	6562
TgSkunk	X	Canada		Skunk	TgSK2
3133	X	USA	CA		4810
3671	X	USA	CA	Sea Otter	3891
3675	X	USA	CA	Sea Otter	4814
ARI	X	USA		Human	SRR350724
WTD-1	X	USA	AL	Deer	6190
TgCatBr5		Brazil			SRR350742
TgCatBr9		Brazil			SRR350728
COUG		Canada	BC		SRR350733
MAS		France	Nice	Human	SRR094469
VAND		French Guiana			SRR094467
B41		USA	PA	Bear	SRR350721
CAST		USA	CA	Human	SRR350735

B



Supplemental Table 2-1: Strain genotyping across the expanded Grigg sequence typing markers
14 sequenced markers across the genomes of 17 strains of Type II, 3 Type I, 2 Type III, 5 Type X, and 8 outgroup strains are characterized by their sequence. All sequences were compared in individual marker maximum likelihood phylogenetic trees. New alleles are determined based on 60% or greater bootstrap support of tree branching. Four representative genomes (I-RH, II-ME49, III-VEG, X-WTD) are utilized to characterize the alleles seen on the phylogenetic trees. **A)** Type II sequence typing markers colored. 17 Type II and the 3 representative genomes (I-RH, II-ME49, III-VEG, X-WTD) are shown here. Colors represent the allele characterization from the phylogenetic trees as described above with I-red, II-green, III-blue, and X-purple. Shades of colors (such as light green versus dark green) differentiate discrete genetic alleles that possess less than 60% bootstrap support differentiation from known alleles. Pink and yellow alleles represent novel alleles found within the Type II strains. **B)** Allele characterizations of all strains used here at all markers. All 35 isolates representing the diversity of *Toxoplasma* strains are characterized at 14 markers across the nuclear and organellar genome. ToxoDB genotype is shown where it matches the observed typing of these strains. Alleles are determined as in A by phylogenetic divergence on maximum likelihood trees. Strains are grouped by typing marker sequencing genotypes.

A

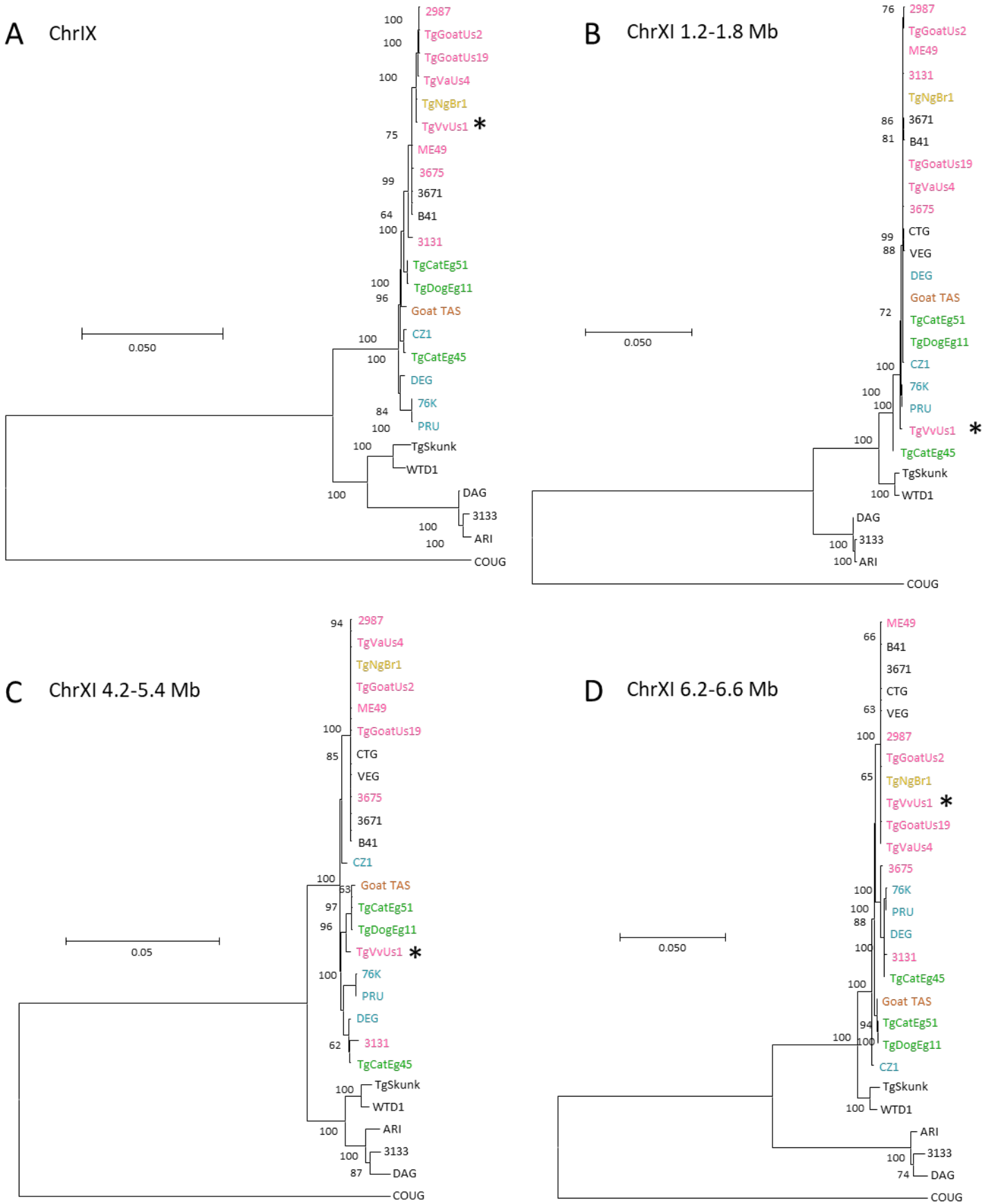
Strains	Region	APICO	c228	c292	L358	PK1	BAG1	SAG1	SRS2	SAG2	Beta Tubulin	GRA6	UPRT	SAG3	EF1
Chromosome		apicoplast	Ib	III	V	VI	VIIb	VIII	VIII	VIII	IX	X	XI	XII	XII
Size		848	657	689	690	1029	855	482	1132	1056	527	846	4037	311	2623
ME49	USA														
RH_88	USA														
VEG	USA														
WTD1	USA														
TgGoatUs2	USA														
TgGoatUs19	USA														
TgNgBr1	Brazil														
3131	USA														
2987	USA														
TgDogEg11	Egypt														
TgVaUs4	USA														
TgVvUs1	USA														
DEG	France														
TgCatEg51	Egypt														
3675	USA														
CZ1	Czech Republic														
Goat_TAS	Australia														
TgCatEg45	Egypt														
76K	France														
PRU	France														

B

Strains	Genotype	APICO	c228	c292	L358	PK1	BAG1	SAG1	SRS2	SAG2	Beta Tubulin	GRA6	UPRT	SAG3	EF1	Strains
Chr		Apicoplast	Ib	III	V	VI	VIIb	VIII	VIII	VIII	IX	X	XI	XII	XII	
Size		848	657	689	690	1029	855	482	1132	1056	527	846	4037	311	2623	
Type II	#1	II	II	II	II	II	II	II	II	II	II	II	II	II	II	ME49
		II	II	II	II	II	II	II	II	II	II	II	III	II	u-1	TgGoatUs2, TgGoatUs19, TgNgBr1
		II	II	II	II	u-1	II	II	II	II	II	II	III	II	u-1	3131
		I	II	II	II	II	II	II	II	II	II	II	III	II	u-1	2987, TgDogEg11, TgVaUs4, TgVvUs1
		I	II	II	II	II	II	II	II	II	II	II	III	u-1	u-1	DEG, TgCatEg51
		I	II	II	u-4	II	II	II	II	II	II	II	III	II	u-1	3675
		I	II	II	II	II	u-4	II	II	II	II	II	III	II	u-1	CZ1
		I	II	II	II	II	u-4	II	u-2	II	II	II	III	II	u-1	Goat_TAS
		I	II	II	II	II	u-4	II	II	II	II	II	u-1	u-1	u-1	TgCatEg45
		I	II	II	II	II	II	II	II	II	II	II	u-1	u-1	u-2	76K, PRU
		I	II	II	X	X	X	X	X	II	X	X	X	II	X	WTD-1
		I	II	II	X	u-4	X	X	X	II	X	X	X	II	X	TgSkunk
		I	II	II	X	II	X	II	II	II	II	II	III	II	X	3671, B41
		I	II	II	u-1	u-2	X	X	X	II	X	u-1	X	II	X	3133, ARI
		I	II	II	u-1	u-3	II	X	X	II	X	u-1	X	II	X	DAG
Type I	#10	I	I	I	I	I	I	I	I	I	I	I	I	I	I	GT1, RH-88, RH-JSR
		I	II	u-1	II	u-5	u-3	I	I	u-1	X	u-2	u-3	u-2	u-3	COUG
		I	u-2	u-3	I	u-6	u-1	I	I	I	u-1	u-5	u-3	III	I	TgCatBr9
		III	u-1	u-2	u-2	I	u-1	u-1	I	I	I	u-3		I	I	CAST
		I	u-2	u-3	u-3	u-6	u-2	I	I	III	u-1	u-4	u-2	III	III	TgCatBr5
		I	III	u-3	u-3	u-7	u-4	u-2	u-3	u-2	u-1	u-4	u-2	III	u-4	MAS
Type III	#2	III	III	III	III	III	II	II	II	III	III	III	III	III	III	CTG, VEG
		u-1	u-3	u-3	u-4	u-8	I	u-3	u-4	u-3	III	u-6	u-4	I	u-5	VAND

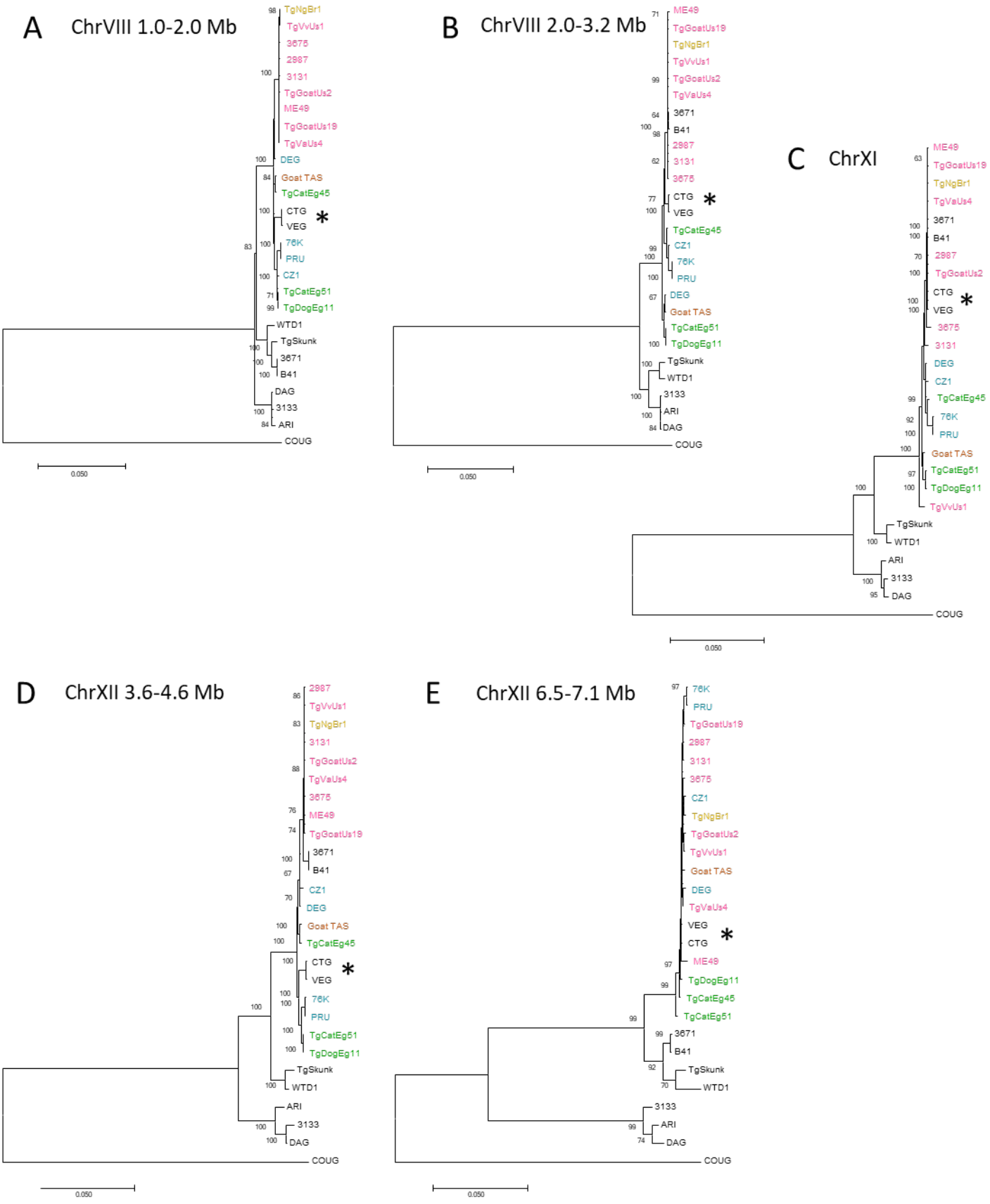
Supplemental Figure 2-2: Phylogenetic trees support the recombination across the genome of TgVvUs1 in Figure 5A

SNPs from the WGS of all strains were extracted in either whole chromosomes or portions of the chromosome as determined by SNP density fingerprinting patterns. Maximum likelihood trees of these SNPs were generated using 100 bootstrap support. Type II strains were colored based on geographical isolation as previously (North America, South America, Eurasia, Africa, Australia). An asterisk denotes the TgVvUs1 strain. **A)** Maximum likelihood tree of all SNPs in chromosome IX. **B)** Maximum likelihood tree of SNPs in chromosome ChrXI from 1.2-1.8 Mbp. **C)** Maximum likelihood tree of SNPs in chromosome ChrXI from 4.2-5.4 Mbp. **D)** Maximum likelihood tree of SNPs in chromosome ChrXI from 6.2-6.6 Mbp.


A
ChrIX
2987
TgGoatUs2
TgGoatUs19
TgVaUs4
TgNgBr1
TgVvUs1 *
ME49
3675
3671
B41
3131
TgCatEg51
TgDogEg11
Goat TAS
CZ1
TgCatEg45
DEG
76K
PRU
TgSkunk
WTD1
DAG
3133
ARI
COUG
0.050
B
ChrXI 1.2-1.8 Mb
2987
TgGoatUs2
ME49
3131
TgNgBr1
3671
B41
TgGoatUs19
TgVaUs4
3675
CTG
VEG
DEG
Goat TAS
TgCatEg51
TgDogEg11
CZ1
76K
PRU
TgVvUs1 *
TgCatEg45
TgSkunk
WTD1
DAG
3133
ARI
COUG
0.050
C
ChrXI 4.2-5.4 Mb
2987
TgVaUs4
TgNgBr1
TgGoatUs2
ME49
TgGoatUs19
CTG
VEG
3675
3671
B41
CZ1
Goat TAS
TgCatEg51
TgDogEg11
TgVvUs1 *
76K
PRU
DEG
3131
TgCatEg45
TgSkunk
WTD1
ARI
3133
DAG
COUG
0.05
D
ChrXI 6.2-6.6 Mb
ME49
B41
3671
CTG
VEG
2987
TgGoatUs2
TgNgBr1
TgVvUs1 *
TgGoatUs19
TgVaUs4
3675
76K
PRU
DEG
3131
TgCatEg45
Goat TAS
TgCatEg51
TgDogEg11
CZ1
TgSkunk
WTD1
ARI
3133
DAG
COUG
0.050

Supplemental Figure 2-3: Phylogenetic trees support the inheritance of different Type II geographically defined haploblocks in the genome of VEG seen in Figure 5B

SNPs from the WGS of all strains were extracted in either whole chromosomes or portions of the chromosome as determined relevant by SNP density fingerprinting patterns. Maximum likelihood trees of these SNPs were generated using 100 bootstrap support. Type II strains were colored based on geographical isolation as previously (North America, South America, Eurasia, Africa, Australia). An asterisk denotes the Type III VEG and CTG strains. **A)** Maximum likelihood tree of SNPs in chromosome ChrVIII from 1.0-2.0 Mbp. **B)** Maximum likelihood tree of SNPs in chromosome ChrVIII from 2.0-3.2 Mbp. **C)** Maximum likelihood tree of all SNPs in chromosome ChrXI. **D)** Maximum likelihood tree of SNPs in chromosome ChrXII from 3.6-4.6 Mbp. **E)** Maximum likelihood tree of SNPs in chromosome ChrXII from 6.5-7.1 Mbp.


A
ChrVIII 1.0-2.0 Mb
TgNgBr1
TgVvUs1
3675
2987
3131
TgGoatUs2
ME49
TgGoatUs19
TgVaUs4
DEG
Goat TAS
TgCatEg45
CTG
VEG
76K
PRU
CZ1
TgCatEg51
TgDogEg11
WTD1
TgSkunk
3671
B41
DAG
3133
ARI
COUG
0.050
B
ChrVIII 2.0-3.2 Mb
ME49
TgGoatUs19
TgNgBr1
TgVvUs1
TgGoatUs2
TgVaUs4
3671
B41
2987
3131
3675
CTG
VEG
TgCatEg45
CZ1
76K
PRU
DEG
Goat TAS
TgCatEg51
TgDogEg11
TgSkunk
WTD1
3133
ARI
DAG
COUG
0.050
C
ChrXI
ME49
TgGoatUs19
TgNgBr1
TgVaUs4
3671
B41
2987
TgGoatUs2
CTG
VEG
3675
3131
DEG
CZ1
TgCatEg45
76K
PRU
Goat TAS
TgCatEg51
TgDogEg11
TgVvUs1
TgSkunk
WTD1
ARI
3133
DAG
COUG
0.050
D
ChrXII 3.6-4.6 Mb
2987
TgVvUs1
TgNgBr1
3131
TgGoatUs2
TgVaUs4
3675
ME49
TgGoatUs19
3671
B41
CZ1
DEG
Goat TAS
TgCatEg45
CTG
VEG
76K
PRU
TgCatEg51
TgDogEg11
TgSkunk
WTD1
ARI
3133
DAG
COUG
0.050
E
ChrXII 6.5-7.1 Mb
76K
PRU
TgGoatUs19
2987
3131
3675
CZ1
TgNgBr1
TgGoatUs2
TgVvUs1
Goat TAS
DEG
TgVaUs4
VEG
CTG
ME49
TgDogEg11
TgCatEg45
TgCatEg51
3671
B41
TgSkunk
WTD1
3133
ARI
DAG
COUG
0.050

Chapter 3 - Virulence Shift in a Sexual Clade of Type X *Toxoplasma* Infecting Southern Sea Otters

Kennard, Andrea[1]; Khan, Asis[1]; Miller, Melissa A.[2]; Miller, Natalie[3]; Quiñones, Mariam[4]; James, Erick R.[5]; Roos, David S.[3]; Conrad, Patricia A.[6]; and Grigg, Michael E.[1,5].

[1]Laboratory of Parasitic Diseases, NIAID, National Institutes of Health, Bethesda, MD. [2]California Department of Fish and Game, Santa Cruz, CA. [3]Department of Biology and Penn Genome Frontiers Institute, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [4]Bioinformatics and Computational Biosciences Branch, NIAID, National Institutes of Health, Bethesda, MD. [5]Department of Medicine, University of British Columbia, Vancouver, BC. [6]School of Veterinary Medicine, University of California, Davis, CA.

Introduction

How successful microbes maintain reservoirs of virulent strains in nature is understudied and is an important paradigm of infectious diseases. While many studies document how microbial pathogens acquire or admix genetic material to influence their pathogenicity, relatively less is known about how emergent strains are selectively maintained in nature. It is well established that viruses utilize reassortment to rapidly produce admixture genomes that possess a range of altered biological potential, including virulence (Twiddy and Holmes 2003; Vijaykrishna et al. 2011). Bacteria transfer mobile plasmid elements via conjugation that encode pathogenicity islands and confer an altered virulence potential on those strains that receive them (Bellanger et al. 2014; Schneider et al. 2011; Trevors 1998; Wardal et al. 2014; Heitman 2010). Among both fungi and protozoa, genetic hybridization by sexual recombination is a potent way to produce genomic variation or new admixture lines that have increased virulence and are

capable of causing disease outbreaks or altering host range (Ni et al. 2011; English, Adomako-Ankomah, and Boyle 2015; Grigg, Bonnefoy, et al. 2001). What is less clear are the mechanisms at play that allow highly prevalent pathogens to maintain virulent strains cryptically in nature. Work shown herein sought to test whether generalist pathogens, such as the protozoan parasite *Toxoplasma gondii*, leverage their broad range of intermediate hosts to selectively partition parasite genotypic diversity and virulence potential.

Toxoplasma gondii is a highly successful and prevalent protozoan pathogen that infects a wide range of wildlife, livestock and 30-80% of humans worldwide (Boothroyd and Grigg 2002; Boyer et al. 2011; Dubey et al. 2011). Its success in nature is largely attributed to its highly flexible life cycle. It can be propagated asexually by carnivory among all warm-blooded vertebrates (via orally infectious tissue cysts), sexually within its definitive felid host by self-mating (when a single parasite undergoes fertilization and sexually clones itself), or by out-crossing (producing as many as 10^8 genetic admixture hybrids that are transmissible as environmentally stable, highly infectious oocysts). Genetic hybridization has been previously shown to produce genotypes that possess a wide spectrum of biological potential, including altered virulence or a capacity to expand as successful epidemic clones (Boyle et al. 2006; Grigg, Bonnefoy, et al. 2001; Wendte, Miller, Lambourn, et al. 2010). Importantly, parasite clones possess a broad spectrum of disease states that are highly dependent on both parasite genetics and the animal species infected (Boothroyd 2009; Boothroyd and Grigg 2002; Howe and Sibley 1995; Lorenzi et al. 2016). In rodents, low dose inoculums of Type I *Toxoplasma* strains are uniformly lethal to laboratory mice (Sibley and Boothroyd 1992) whereas Type II strains are considerably less virulent and routinely establish chronic, transmissible infections (Howe and Sibley 1995). In contrast, wild-derived CIM mice that encode resistant alleles of immunity-

related GTPases (IRGs) confer resistance to virulent Type I strains by inactivating parasite-specific secreted kinases (ROP18/ROP5) that control mouse susceptibility to infection (Lilue et al. 2013; Steinfeldt et al. 2010; Gazzinelli et al. 2014). The end result of this host-parasite genetic interplay determines the potential for a disease-producing virulent clone to expand or be maintained cryptically within a strain-specific, or host species-specific, manner, but this has not yet been demonstrated in natural populations and has only been inferred from laboratory studies. Likewise, laboratory rats are resistant to Type I infections and fail to transmit parasites, but again this is entirely rat species-dependent (Cavailles et al. 2006). The molecular basis for this resistance is an allele-dependent activation of the NLRP1 inflammasome resulting in macrophage pyroptosis and the inhibition of parasite growth in Lewis (LEW) and Sprague Dawley (SD) rats, but not in Brown Norway (BN) or Fischer CDF rats, which go on to establish chronic, transmissible infections of mouse-virulent Type I strains (Cirelli et al. 2014; Ewald, Chavarria-Smith, and Boothroyd 2014; Gorfu et al. 2014). While these laboratory studies suggest the possibility that virulent strains can be maintained cryptically across the pathogen's host range (Dubey et al. 1999), this phenomenon has not been systematically investigated nor demonstrated to occur in a natural setting.

Between 1998 and 2004, *Toxoplasma gondii* emerged as a significant disease in California sea otters (*Enhydra lutris nereis*), a federally listed threatened species (Conrad et al. 2005; Miller et al. 2004). Full necropsies performed on each stranded sea otter identified a wide spectrum of disease sequelae among *Toxoplasma*-infected otters, ranging from chronic asymptomatic infection to fatal meningoencephalitis at time of death (Miller et al. 2004). Importantly, the majority of sea otters were found to be infected by a single outbreak clone, referred to as Type X (or HG12) (Conrad et al. 2005). Based on multi-locus PCR-DNA

sequencing (MLST) using a limited set of genotyping markers, Type X was defined as the fourth clonal lineage in the US (Khan, Dubey, et al. 2011) and was shown to be highly prevalent in sylvatic niches as chronic, subclinical, or mild infections in 47% of US wildlife (Dubey et al. 2011; VanWormer et al. 2014). The range of disease sequelae detected in sea otters, however, is not consistent with infection by a single clone. Other possibilities proposed to explain variation in disease susceptibility have included co-infection (Gibson et al. 2011), exposure to environmental pollutants, or toxins that cause immunosuppression (Kreuder et al. 2003). More recently, WGS performed on *Toxoplasma* isolates has called into question whether MLST analysis is sufficiently resolved to predict clonotypes because it fails to capture extant genetic heterogeneity (Lorenzi et al. 2016). Indeed, a previous study concluded that Type X is comprised of at least two genotypes (Sundar et al. 2008), suggesting that strain variation may likewise play a contributing role in sea otter disease in this novel host-parasite interaction.

To test for this and to determine the true extent of genetic diversity among naturally circulating Type X isolates, a wide selection of linked and unlinked markers were sequenced here across the nuclear and organellar genomes for 53 isolates collected longitudinally from infected sea otters across a seven-year period. Based on these results, we selected 16 isolates for WGS. Analysis of our data support a model whereby Type X exists as a recombinant clade of strains that resemble F1 progeny from a natural cross between a Type II strain and a novel genotype that has not been identified previously in nature. One sea otter genotype was shown to be widely distributed and had expanded to cause the majority of infections in sea otters, most of which were subclinical and consistent with a strain that has been naturally selected in this mustelid host. This genotype, however, was highly pathogenic to laboratory mice and a population level, forward genetic approach using WGS mapped the mouse virulence gene to the

novel parasite-specific secreted kinase ROP33. This natural population dataset argues that *Toxoplasma* leverages its broad range of intermediate hosts to partition genetic diversity. Further, that intermediate hosts play a central role in the natural selection, expansion and maintenance of cryptically virulent strains in a host species-specific manner and demonstrated that these infected hosts act as reservoirs for pathogenic or epidemic disease in another host species.

Results

Genetic Characterization of *Toxoplasma gondii* Strains Isolated from Sea Otters

In the late 1990s, an increasing number of sea otters stranded along the California coast with protozoal encephalitis due to *Toxoplasma gondii* (Miller et al. 2004; Miller et al. 2008; Sundar et al. 2008; VanWormer et al. 2014). An initial genetic analysis performed on a subset of 35 *T. gondii* isolates established that 21 were a new genetic type, referred to as Type X, but this conclusion was limited to sequence typing at a single locus, GRA6 (Miller et al. 2004). Here was developed a multi-locus (MLST) PCR-DNA sequence-based typing scheme using an additional 4 unlinked markers (BSR4, BAG1, SAG3, and ROP1) that possess a wide-range of phylogenetic strength, including markers under neutral (BSR4), diversifying (GRA6, SAG3, ROP1), and purifying (BAG1) selection to determine the molecular genotypes of 53 *T. gondii* isolates collected longitudinally over a seven-year period between 1998-2004 from fresh, beach-cast California sea otters. Isolates were obtained throughout the California sea otter range (Supplemental Figure 1). Antibody titers against *Toxoplasma* (when available), stranding location, co-infection status with other protozoal agents, and whether brain inflammation was associated with infection was recorded for 52 of the sea otters (Supplemental Figure 1).

Point-source Outbreak of Type II Strains Infecting Sea Otters

To determine the *T. gondii* genotypes among the sea otter isolates, the 5 locus MLST was applied. Typing at just the GRA6 marker identified three sequence types, either a Type II allele in 16 (30%) isolates, or two non-archetypal alleles, previously identified as X or A (Miller et al. 2004; Sundar et al. 2008), in 25 (47%) and 12 (23%) isolates, respectively. Expanding the genetic analysis across five markers showed that 2 of 16 isolates with a Type II allele at GRA6 were predicted to be recombinant strains, because they possessed non-archetypal alleles at ROP1 (3675) or BSR4 and BAG1 (3671), confirming that these strains are not Type II (Figure 1A). Of the 14 otters infected with Type II strains, 12 (86%) were identified within a restricted geographic range, a 40 mile stretch off the California coast line containing Moss Landing and Monterey Bay, which is consistent with a point source outbreak of a *Toxoplasma* strain that is found most frequently infecting human and domestic livestock in the USA. The other two Type II infected otters were recovered from Pismo Beach in Central California. Further, the majority of Type II infected otters (9/14; 64%) were collected during the first half of the study (Supplemental Figure 1), and only males were predominantly infected with this genotype. [Fisher's two-sided exact test II vs. all; p=0.008].

Type X is a Recombinant Clade of Strains

Of the remaining 39 (74%) *Toxoplasma* isolates, collectively referred to as Type X or A by the allele present at GRA6, a maximum of only three alleles were identified at the additional sequenced loci, either a canonical Type II allele or one of two genetically distinct alleles, referred to as "γ" or "δ", that appeared to segregate independently across the isolates (Figure 1A). In total, 8 distinct haplotypes designated A-H were resolved based on their differential inheritance of the limited alleles, consistent with recombination between a Type II strain and a mosaic of two distinct ancestries as the most plausible explanation for the genetic relationship among the non-Type II strains. The Type X haplotype designated A was dominant (19/39; 49%). It was widely distributed across the entire geographic range of the Southern sea otter (Supplemental Figure 1), and it was isolated in every year except 2000. Importantly, this genotype expanded during the time frame of the study period and was responsible for the majority of infections by 2003 (Figure 1B).

Acute Virulence in Mice is Dependent on Type X Haplotype

Mouse virulence is highly dependent on *Toxoplasma* genotype (Grigg, Bonnefoy, et al. 2001; Saeij, Boyle, and Boothroyd 2005; Sibley and Boothroyd 1992). To further characterize the otter isolates, 18 strains, including two Type II and at least one isolate representing each of the 8 distinct Type X haplotypes, were tested for their mouse virulence phenotype. Isolates were selected based largely by their ability to expand *in vitro* in human foreskin fibroblasts. Virulence was assayed in CD-1 outbred female mice using a low-dose, 50 tachyzoite, intra-peritoneal injection model. At this inoculum, mouse avirulent strains (*e.g.* Type II) establish chronic,

transmissible infections whereas mouse virulent strains (*e.g.* Type I) die acutely within 10-14 days.

Among the otter isolates, three mouse virulence phenotypes were identified: virulent (red; all mice died acutely), intermediate virulent (blue; some mice survived acute disease), and avirulent (green; all mice survived acute infection and produced chronic disease). The two otter isolates that possessed a Type II MLST were avirulent (Figure 1C; green) and phenocopied the infection kinetics of two well studied Type II strains, ME49 and 76K. Among the Type X haplotypes, E and G were also avirulent. Three otter isolates from haplotype A, the genotype that had expanded to cause the majority of sea otter infections, were highly pathogenic, and all infected mice died acutely (Figure 1C; red). Haplotypes D and H were likewise highly virulent in mice. The three haplotype F otter isolates all possessed an intermediate virulence phenotype (Figure 1C; blue). Of particular note, the 5 isolates from haplotypes B and C were either avirulent or possessed an intermediate mouse virulence phenotype (Figure 1C). Because the Type X isolates displayed a range of virulence phenotypes, the data collectively support a model whereby Type X represents a clade of recombinants from a genetic admixture of a mouse avirulent Type II strain with at least one other parent that is mouse virulent, rather than an expansion of a clonal lineage. Furthermore, not all isolates within each subgroup displayed equivalent virulence kinetics, in particular isolates from the B and C haplotypes were either avirulent or possessed intermediate virulence, suggesting that the 5 marker MLST was insufficiently resolved to fully capture total genetic and phenotypic diversity among the isolates infecting Southern sea otters.

To explore this possibility, the 18 isolates assayed through mice plus 3 additional isolates, one each from haplotypes A, C and D, were further genotyped at 18 markers (17

nuclear-encoded single copy gene loci plus one microsatellite marker) representing an expanded set of linked and unlinked loci across the genome. Additionally, 2 markers, one each from the organellar genomes of the apicoplast and mitochondria, were included to determine the ancestry of each maternally inherited genome. This represents a significant increase in the number of markers that were previously used to conclude that Type X (also referred to as HG12) is a clonal lineage (Khan, Dubey, et al. 2011; Dubey et al. 2011; Su et al. 2012). Altogether, the 20 marker MLST method provided 15,404 bp of sequence information to determine the genetic relationship among *Toxoplasma* strains isolated from Southern sea otters (Supplemental Table 1).

Type X is Comprised of Twelve Distinct Haplotypes by an Expanded MLST Analysis

To establish the number of haplotypes among the 21 isolates, sequence data for the 17 nuclear-encoded single copy gene loci was concatenated and analyzed using eBURST (Figure 1D). eBURST resolves strains into clonal complexes (CC) that are in linkage disequilibrium. In this case, coded MLST data was used to identify strains that differ at one locus (or share 16 out of 17 markers). Isolates that differed by a single genetic locus were grouped together and connected by a line within each clonal complex. As expected, eBURST clustered the two Type II isolates (2987, 3131) into a single clonal complex with ME49 (CC1). Type X subgroup G isolate 3675, originally identified by its allele at the microsatellite marker ROP1, was included within CC1 because it differed from Type II strains at only SAG4. For the remaining 18 Type X isolates, 15 clustered into 3 clonal complexes (designated CC2-CC4), and 3 unique haplotypes were resolved. This eBURST analysis failed to support a clonal lineage designation for Type X. Further, the increased resolution provided by the 17 markers subdivided the 8 distinct haplotypes defined using 5 loci in Figure 1A into 12 haplotypes (Figure 1D). One isolate each from

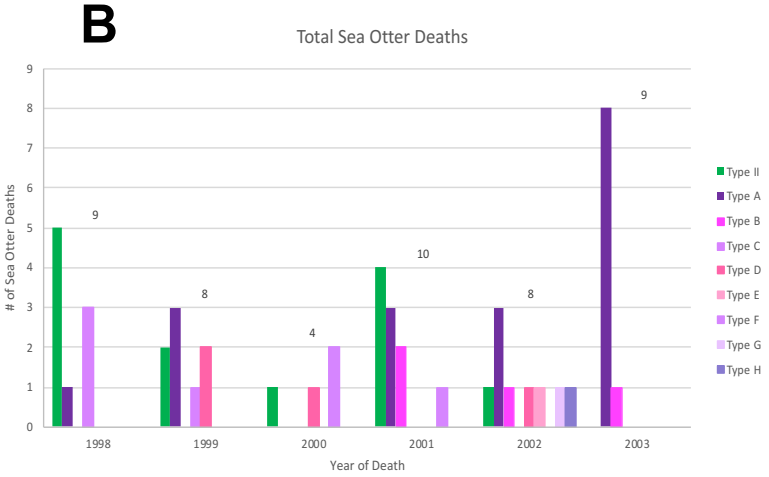
subgroups A (n=4; 3168), C (n=4; 3160), D (n=3; 4166) and F (n=3; 3387) was further resolved as a unique haplotype based on inheritance of alleles at PK1, SAG4 and GRA7 (Supplemental Table 1).

Figure 3-1: Parasite diversity and mouse virulence within the Type X clade
A) 53 isolates genotyped using 5 unlinked markers identified 9 distinct T. gondii genotypes infecting southern sea otters. Type I, II, and III alleles are represented by red, green, and blue, respectively. Sea otters were infected with T. gondii strains that possessed either a non-archetypal allele (purple or orange) or a Type II allele (green) at the loci examined. The number of sea otter infections (#SO) and percentage of total infections (%SO) is shown for each genotype group. Each group was comprised of the following isolates: A (3097, 3142, 3168, 3265, 3483, 3488, 3520, 3637, 3659, 3744, 3786, 3821, 3865, 3947, 3950, 4003, 4045, 4071, 4151), B (3458, 3523, 3728, 3897), C (3026, 3045, 3077, 3160, 4167), D (3133, 3178, 3183, 3451, 4166), E (3819), F (3387, 3429, 3503), G (3675), H (3671), II (2987, 2994, 3005, 3009, 3087, 3131, 3208, 3396, 3521, 3576, 3587, 3636, 3739, 4181). **B)** Total Sea Otter Deaths: number of sea otter deaths between 1998-2003 are charted according to their 5-marker type designation of Type II (green) or sub-types of Type X (range of purples). Type II infections not only clustered geographically (not shown) but also temporally while Type X infections are steady over time. Additionally, the Type X, sub-type A infections which were predominantly avirulent in sea otters show increased sea otter infections over time. **C)** CD1 mice were infected with 50 tachyzoites intraperitoneally with one of 18 isolates representing all singletons (E, G, H) and at least two isolates from each of the other 6 distinct genotypes (II, A, B, C, D, F) recovered from sea otters. Mouse seroconversion and survival was monitored for 30 days. At least two independent infection experiments were performed using 5 mice each for every strain tested. Results shown are for those mice that seroconverted or died acutely during infection. Strains are grouped and colored based on their genotype and virulence in mice. Red, mouse virulent, Blue, intermediate mouse virulent (some mice survived acute infection); Green, mouse avirulent (all mice survived acute infection). Number of mice infected by each genotype and their relative virulence is indicated. **D)** eBURST analysis to identify linkage disequilibrium across 17 nuclear-encoded linked and unlinked markers identified 4 clonal complexes and 3 unique Type X genotypes among 21 isolates selected from the 9 distinct clades in Figure 1A. Isolate colors represent the lineage Type from A: red (Type I), green (Type II), blue (Type III), and purple (Type X; all subgroups (A-H)). Dot size is proportional to the number of isolates with that genotype; larger dots are multi-isolate genotypes. Clonal complexes (CC) are indicated by lines connecting isolates and are highlighted with ovals corresponding to their allelic identity (X-purple, II-green)

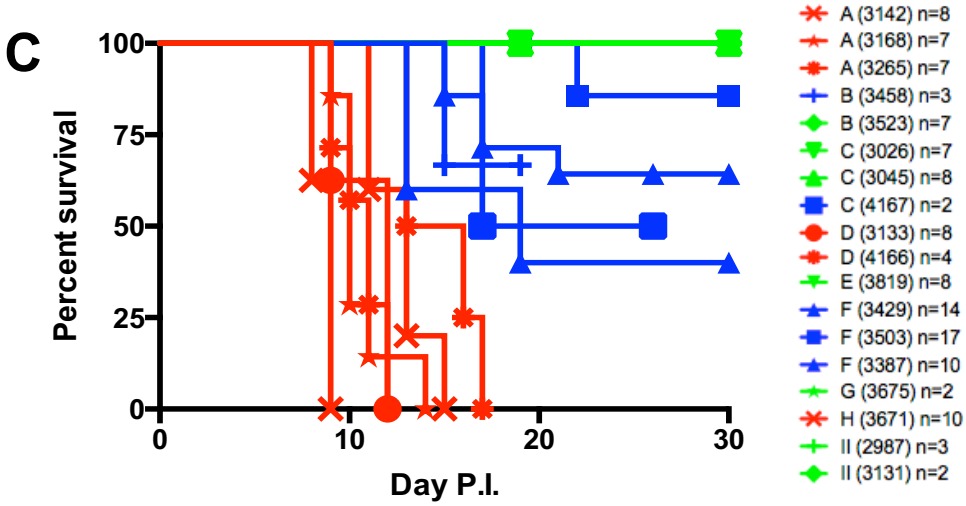
A

Marker	BSR4	BAG1	GRA6	ROP1	SAG3	# SO	% SO
Marker Type	Seq	Seq	Seq	RFLP	Seq		
Chromosome	IV	VIIb	X	XI	XII		
Length	1138	1917	758	Dde1	253		
Type I							
Type II							
Type III							
II						14	26
A						19	36
B						4	7.5
C						5	9.4
D						5	9.4
E						1	1.9
F						3	5.7
G						1	1.9
H						1	1.9

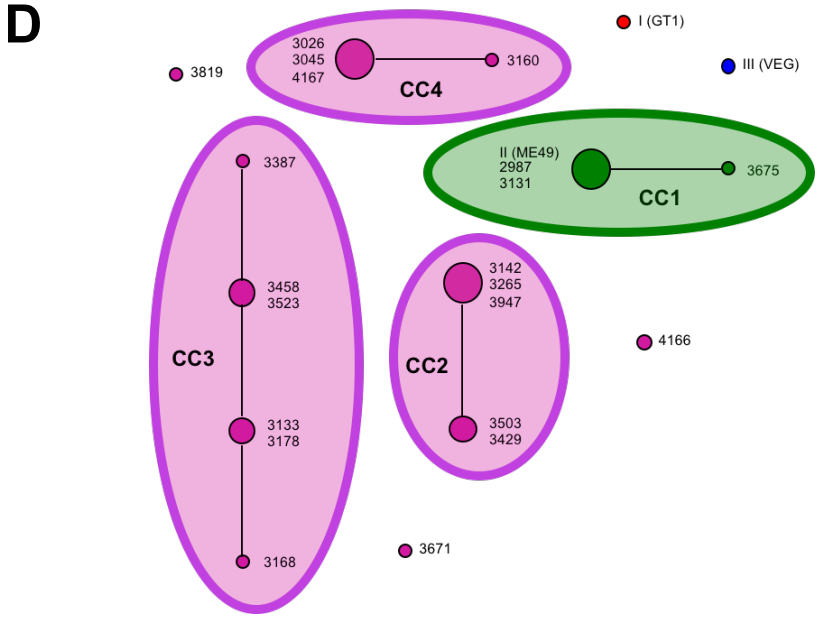
B



C



D



Type X Resembles a Recombinant Clade of F1 Progeny from a Natural Cross

eBURST, however, is insufficiently resolved to distinguish among isolates that are either variants within a clonal lineage but possess alleles that differ by only minor mutational drift or exist as admixtures that possess genetically distinct alleles derived from different parental types. Previous studies suggested that Type X (or HG12) is a clonal lineage infecting sea otters and wildlife in North America, and that HG12 represents an expanded clone from a sexual cross between a Type II strain and a new genetic lineage, that was referred to as "γ", but this was based largely on typing at just two loci, GRA6 or GRA7 (Conrad et al. 2005; Khan, Dubey, et al. 2011; Miller et al. 2004). To investigate if the genetic relationship among the 12 distinct haplotypes is supported by a model of genetic drift, or that of recombination, individual maximum likelihood trees were created for each of the 17 nuclear-encoded single gene copy loci, the microsatellite locus, and the two organellar loci (Supplemental Figure 1). To differentiate between alleles undergoing minor mutational drift, from those that have evolved independently as distinct genetic outgroups, 1000 bootstrap replicates were run for each tree, and supported nodes above 60% were indicated at each marker. For each tree, the 19 isolates identified as Type X were labeled in purple and isolates belonging to the clonal lineage I (n=1), II (n=3), and III (n=1) were labeled in red, green, and blue respectively (Figure 2).

At 5 genetic markers located on chromosomes Ib, II, III, XI, and XII, all 12 distinct Type X haplotypes possessed a canonical Type II allele (green), with no minor mutational drift detected, consistent with Type II being one of the parental genetic backgrounds for the Type X clade (Supplemental Table 1). At the L358 marker on chromosome V, however, all Type X isolates except 3675 possessed a Type I or "α" allele (red), indicating the presence of mixed ancestry. Further, at the BSR4 marker on chromosome IV, all Type X isolates except 3675

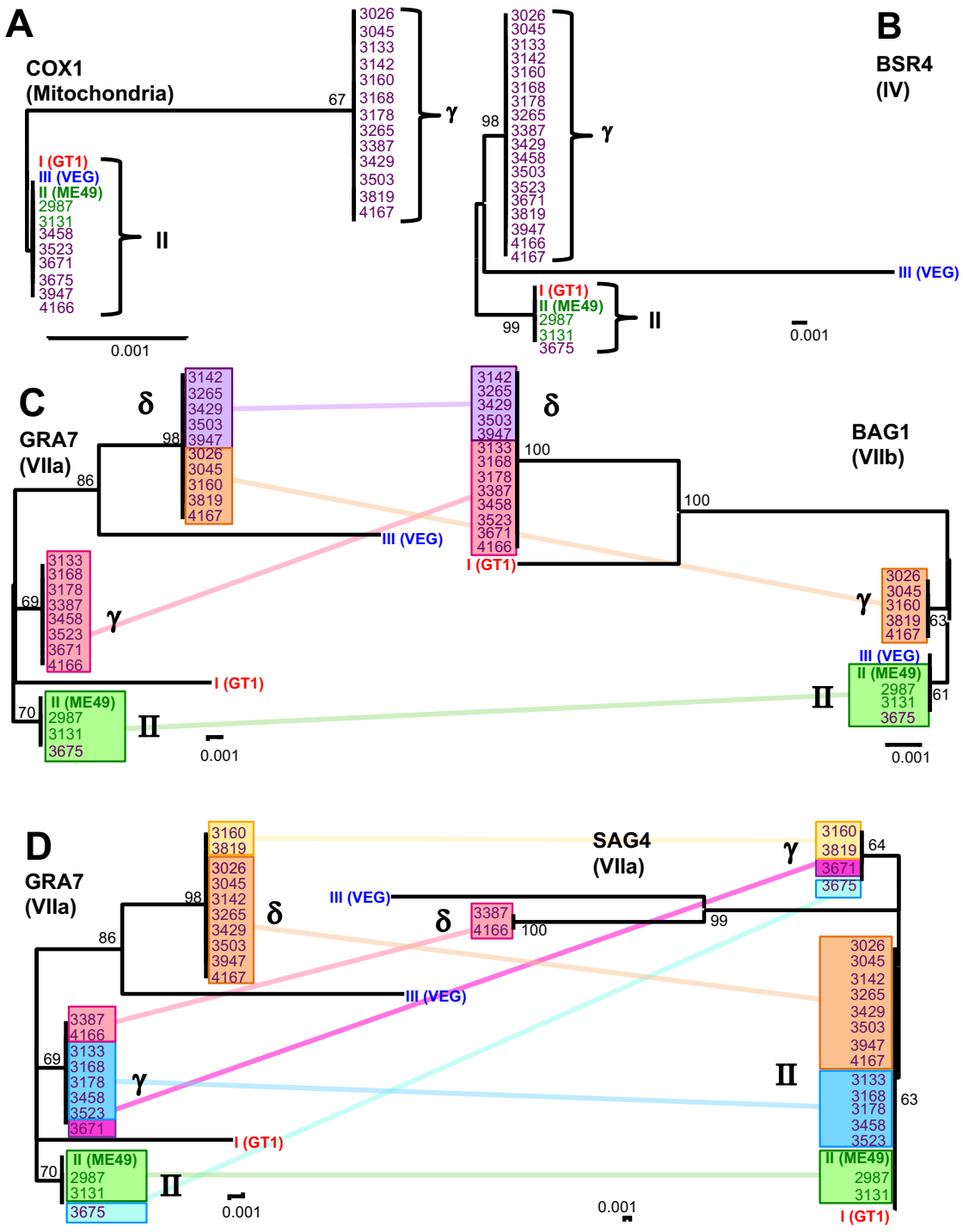
possessed an entirely novel allele that was readily differentiated from Type I, II and III strains with strong bootstrap support (Figure 2B). At all remaining loci, Type X isolates possessed either a Type II or one of 2 genetically distinct alleles, which could not be explained by minor mutational drift. Hence, across the 17 nuclear-encoded markers, each isolate possessed either a Type II allele, or an allele of distinct ancestry that was referred to as belonging to the "α" (red), "γ" (purple) or "δ" (orange) lineages.

The maternally inherited organellar genome markers were likewise mixed. At the mitochondrial marker COX1, two alleles were identified, either an allele common to Type I, II and III strains or a novel allele designated "γ" (Figure 2A). Further, 3 distinct alleles were identified at the apicoplast marker APICO: A Type I or "α" allele (red), a Type II allele (green) or a novel "γ" allele (purple), supporting the mixed ancestry designation (Figure 2B). Each of the 12 Type X haplotypes was therefore a mosaic of mixed ancestry that possessed some combination of Type II, Type I, or "γ" alleles that had segregated independently across the loci investigated. The data are consistent with a sexual cross model in which Type X resembles a recombinant clade of F1 progeny that was recently derived, without sufficient time to develop minor mutational drift (Supplemental Table 1).

In support of the sexual cross model, incongruity between phylogenies for each haplotype was readily observed between markers located in different portions of the genome. In Figure 2C, 8 isolates shared the same ancestry between the two genetic loci located on separate chromosomes VIIa and VIIb (GRA7 and BAG1 respectively) and were shaded either green (Type II alleles at both loci) or purple (γ lineage alleles at both loci). In contrast, 13 isolates had independently segregated chromosomes of mixed ancestry, colored orange and pink (to depict isolates that were discordant and possessed different ancestral alleles at each locus). The crossing

of orange and pink lines highlighted this lack of incongruence (Figure 2C). For example, isolate 3671, had a γ lineage allele at GRA7 but a δ lineage allele at BAG1. Furthermore, recombination within a chromosome was similarly observed at linked markers for 18 strains (Figure 2D). For example, isolate 3387, had a γ lineage allele at GRA7 but a δ lineage allele at SAG4, whereas 3675 had an II lineage allele GRA7 but a γ lineage allele SAG4, formally establishing that recombination has occurred and supporting genetic hybridization as the most likely explanation for the origin of the 12 Type X haplotypes.

Figure 3-2: Phylogenic biallelism and allelic segregation between sequenced markers
Maximum likelihood trees of sequenced markers. Isolates are colored based on their genotype from Figure 1A (I: red, II: green, III: blue, X: purple). **A)** Mitochondrial marker, COX1. **B)** Genomic loci BSR4. New alleles are classified by differentiation of greater than 60% bootstrap support. The γ lineage is significantly different than Type II alleles and, on some markers, can also be distinguished from a second unique clade only observed in Type X isolates, dubbed δ. **C)** Comparison of unlinked genomic markers GRA7 and BAG1. Chromosomal segregation between groups of isolates are highlighted. Isolates of δ lineage that lack differential segregation are highlighted in purple, those that retain Type II lineage are green. Isolates that show lineage recombination via chromosomal segregation between II and γ/δ alleles are highlighted pink and orange depending on allelic combination. **D)** Comparison of linked genomic markers GRA7 and SAG4. Allelic recombination is highlighted within this chromosome. Isolates retain Type II lineage alleles are highlighted in green. Isolates that show recombination between the γ/δ lineage alleles and the Type II lineage alleles are highlighted in pink and orange depending on recombination directionality.


A
COX1
(Mitochondria)
67
3026
3045
3133
3142
3160
3168
3178
3265
3387
3429
3503
3819
4167
γ
I (GT1)
III (VEG)
II (ME49)
2987
3131
3458
3523
3671
3675
3947
4166
II
0.001
B
BSR4
(IV)
98
3026
3045
3133
3142
3160
3168
3178
3265
3387
3429
3458
3503
3523
3671
3819
3947
4166
4167
γ
III (VEG)
99
I (GT1)
II (ME49)
2987
3131
3675
II
0.001
C
GRA7
(VIIa)
δ
98
86
3142
3265
3429
3503
3947
3026
3045
3160
3819
4167
III (VEG)
69
3133
3168
3178
3387
3458
3523
3671
4166
γ
I (GT1)
70
II (ME49)
2987
3131
3675
II
0.001
100
100
BAG1
(VIIb)
63
61
0.001
D
GRA7
(VIIa)
98
86
3160
3819
3026
3045
3142
3265
3429
3503
3947
4167
δ
III (VEG)
69
3387
4166
3133
3168
3178
3458
3523
3671
γ
I (GT1)
70
II (ME49)
2987
3131
3675
II
0.001
δ
3387
4166
100
99
SAG4
(VIIa)
γ
3160
3819
64
3671
3675
3026
3045
3142
3265
3429
3503
3947
4167
II
3133
3168
3178
3458
3523
63
II (ME49)
2987
3131
I (GT1)
0.001

Type X is an Admixture Cross Between Type II and a Novel Genetic Background

A subset of Type X isolates, 2 each from haplotypes A (3142, 3265), C (3026, 3045), F (3503, 3429) and one each from D (3178) and H (3671) were hybridized against a photolithographic microarray that possessed 1517 polymorphic Type I, II and III strain-specific genotyping probes distributed genome-wide (Khan, Miller, et al. 2011). This was done to test whether the MLST genetic markers accurately predicted the presence of large genome-wide haplotype specific blocks, consistent with genetic hybridization. Each Type-specific SNP on the CGH array is represented as a dot colored either red, green or blue respectively for a hybridizing strain that possesses a Type I, II or III SNP at the probe position. Dots colored grey identify probes that fail to hybridize, consistent with a strain that does not possess a Type I, II or III specific SNP at the probe position. Hybridization with DNA from a Type I (GT1), Type II (ME49), and Type III (CTG) strain identified hybridization patterns consistent with genotype, indicating that all genotype-specific probes were functioning as expected (Figure 3A). Hybridization with the Type X isolates identified large, contiguous Type II haploblocks, as evidenced by hybridization of all Type II-specific probes present on Chr III, VI, VIIb and XII, for example. However, in other portions of the genome, the SNP diversity pattern was novel. The hybridization pattern in these regions either shared patch-work similarity with Type I (i.e., at Chr Ib, VI) or was highly divergent with large contiguous blocks of SNP probes colored grey that failed to hybridize, which was consistent with introgression of a non-Type I, II or III genetic ancestry, rather than minor mutational drift as the explanation for lack of reactivity (i.e., Chr IV, XI, right end of Chr VII, XII, left end of Chr VIII).

The pattern of hybridizing SNPs was unique for each of the seven Type X isolates examined, but it was only minimally different; either 3, 5, or 15 SNP differences were detected

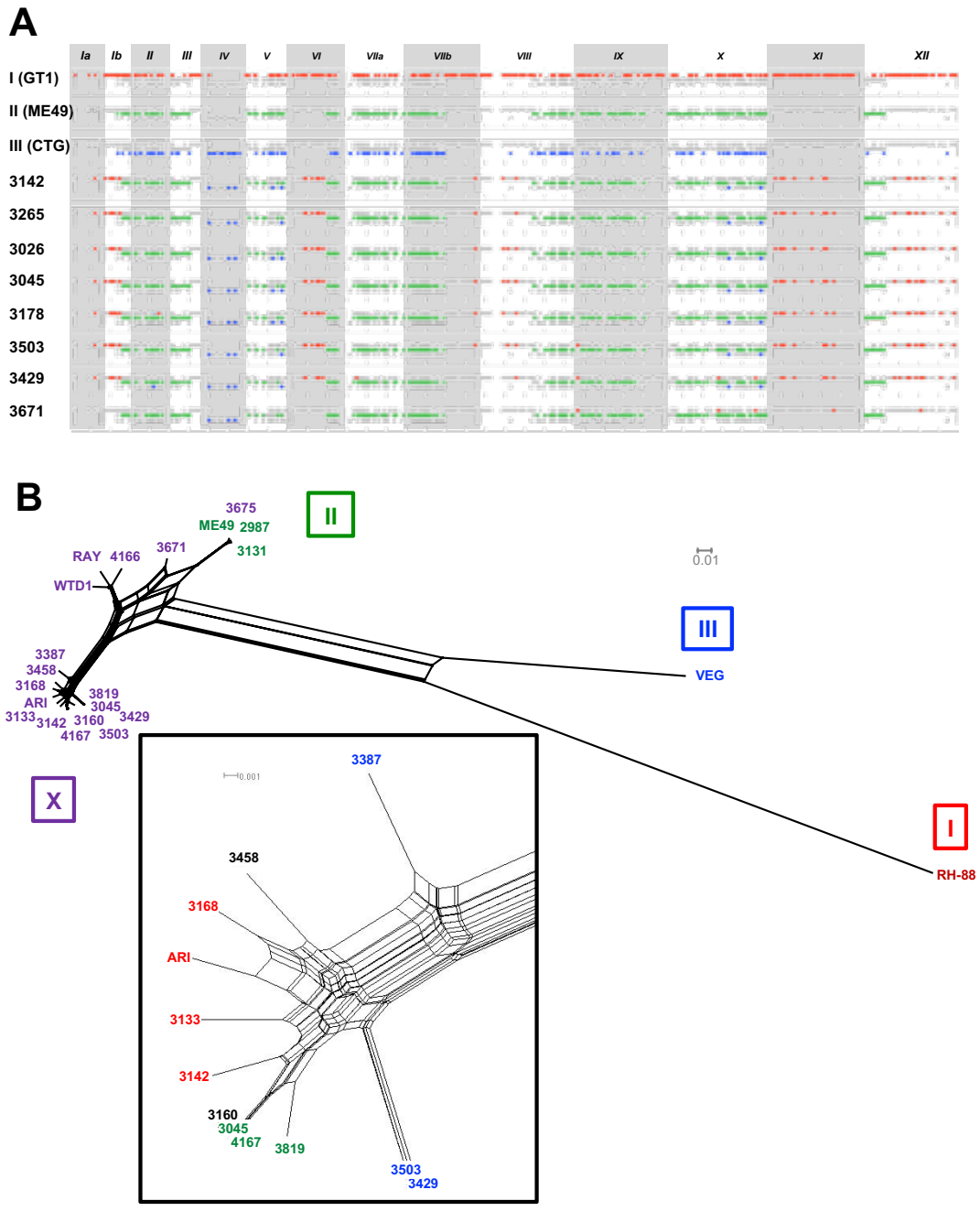
in pairwise comparisons within haplotype A (3142, 3265), C (3026, 3045), and F (3503, 3429), respectively (Figure 3A). At this resolution, it was not possible to distinguish between CGH array hybridization efficiency, from that of minor mutational drift. Indeed, the CGH arrays each had approximately the same number of SNP differences (3-15) between haplotypes as they did within a haplotype, which is more consistent with hybridization efficiency as the probable explanation for these hybridization differences. It did, however, establish that 3671 is a distinct genetic admixture because it possessed a different Type II hybridization pattern at Chr Ia, left hand of Chr Ib, VI, XI and right side of Chr XII, which was different from all other Type X strains analyzed (Figure 3A).

Type X is a Recombinant Clade at WGS Resolution

The lack of resolution using the CGH array approach indicated that WGS is required to infer an accurate genetic history model for the Type X strains. Using genome-wide polymorphism data derived by identifying all variant SNP positions after reference mapping a Type I, II, III, 16 Type X, and three previously WGS sequenced HG12 isolates (WTD-1, RAY, ARI) against the published ME49 genome, an unrooted NeighborNet tree was generated (Figure 3B). The NeighborNet analysis established that all Type X isolates, similar to Type I and III strains, shared definitive genome-wide ancestry with Type II, represented by the reticulated pattern of edge blocks that depict recombination events between Type II and the different genetic backgrounds within X, I, and III. However, NeighborNet analysis did not support the clonal lineage designation as the majority of Type X isolates appeared on separate branches (Figure 3B, Inset). Only clonal complexes 1 and 4 (CC1, Type II; CC4, 3045, 3160, 4167) were supported at WGS resolution, and 3 previously sequenced HG12 isolates from 2 people (ARI, RAY) and a

deer (WTD-1) formed well-supported clades with Type X isolates 4166 (WTD-1; RAY) and 3168 (ARI) indicating that Type X has a wider distribution beyond sea otter infection. Outside of the reticulated network of edge blocks, the branch length for each Type X isolate was significantly less than that observed for Type I and III, synonymous with a more recent origin, without sufficient time to accumulate private SNPs by mutational drift. When murine virulence data for each genotype was mapped onto the NeighborNet tree (see inset, Figure 3B), a clear partitioning of the different virulence phenotypes was resolved along branches within the network, consistent with Type X isolates existing as sister progeny with only minor mutational drift detected at each supported branch.

Figure 3-3: Genome-wide SNP typing of Type X displays haploblock recombination across the genome
A) Microarray hybridization of Type X strains to clonal characterization probes across the genome of Type X isolates. SNPs are represented by a dot in one of three rows indicating where the isolate has hybridized to the microarray with hybridization characteristic of lineage Type I (red), II (green), or III (blue). Grey indicates no hybridization at this location. Chromosomes are represented as alternating grey and white bars. **B)** Type X linkage disequilibrium and network reticulation demonstrates recombination across the genomes of Type X. NeighborNet tree based on whole genome sequence identified 568,592 SNP variant positions across the 19 Type X strains reference mapped to ME49. Type I (red), II (green), III (blue), and X (purple) strains annotated in the box beside the group was based on the previous MLST designation. Two clusters of Type X strains were identified. Strains in inset are colored based on their murine virulence: virulent (red), intermediate virulent (blue), and avirulent (green).



To identify the size, genome distribution, and total number of Type II admixed haploblocks introgressed into each Type X isolate, pairwise SNP diversity plots were generated for all Type X isolates, as well as reference Type I and III strains that are known to have recombined with Type II. The Type I GT1 strain possessed an ME49 Chr Ia and IV, and an ME49 admixture

block at the left side of Chr VIIa, and right end of Chr XI (Figure 4). As expected, the Type III VEG strain possessed many more ME49 admixture blocks distributed genome-wide than GT1, comprising ~40% of its genome (Minot et al. 2012). Only Type X haplotype H (3671) possessed haploblocks that were highly similar in sequence to ME49, on Chr Ia, II, III, VI, VIII, IX, XI, and XII (Figure 4). In other regions, however, 3671 shared regions clearly ME49-like, but that appeared to have diverged somewhat by mutational drift (average 3-5 SNPs per 10kb block) on Chr II or introgressed into Chr IV, V, VIIa, VIIb, VIII, IX, and XII. Intermixed within the Type II regions of the 3671 genome were a limited number of large haploblocks containing divergent SNP density synonymous with hybridization by the distinct γ or δ genetic ancestry (average of 50-100 SNPs per 10kb block). When the pairwise SNP density analysis was expanded to different Type X haplotypes, isolates within each haplotype possessed highly similar patchwork mosaics of Type II-like or divergent (Type X) haploblocks that were specific to each haplotype. For example, haplotype A isolates 3142 and 3168 SNP density plots were highly similar to each other, and to haplotype C isolate 3045, but were readily distinguishable from isolates 4166 and 3503, strains within haplotypes D and F, respectively (Figure 4). Specifically, isolate 3142 had a Type X haploblock inheritance pattern at the right end of Chr V, and in the middle of Chr Ib, VIIa, VIII, IX and XII whereas 4166, which was indistinguishable from a previously sequenced HG12 strain (RAY) recovered from a human patient, and possessed either Type II or 3671 haploblocks in these regions (Figure 4). Although the haplotype F isolate 3503 was highly similar in genomic organization to 3142, it possessed a Type II haploblock at the right end of Chr V that readily distinguished it from haplotype A. The pairwise SNP analysis established that the majority of Type X haplotypes possessed a genomic architecture that is strikingly similar, but different at a limited number of admixture blocks. Coupled with the low allelic diversity, the data

support a genetic history model whereby Type X resembles a sexual clade of recently-derived natural recombinants from a relatively limited number of crosses.

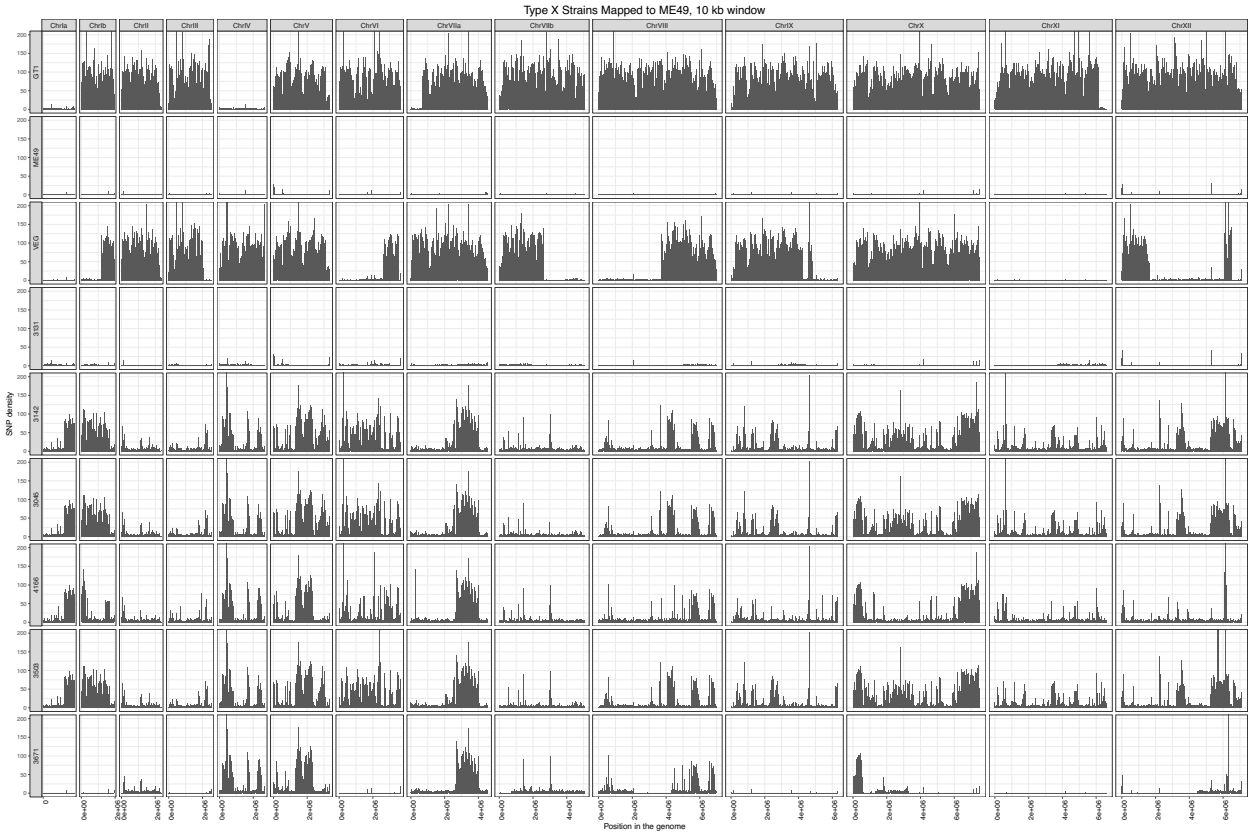


Figure 3-4: Whole genome SNP density displays clear recombination of haploblocks of II and γ ancestry within the Type X strains.
Isolates are reference mapped to the ME49 Type II strain and SNPs per 100 kbp window were graphed. Areas of higher SNP density are least like the Type II strain in that window. SNP density shows clear divisions of haploblocks of Type II and divergent SNP density haploblocks.

PopNet Analysis Identifies Only Limited Admixture Blocks Among Type X Isolates

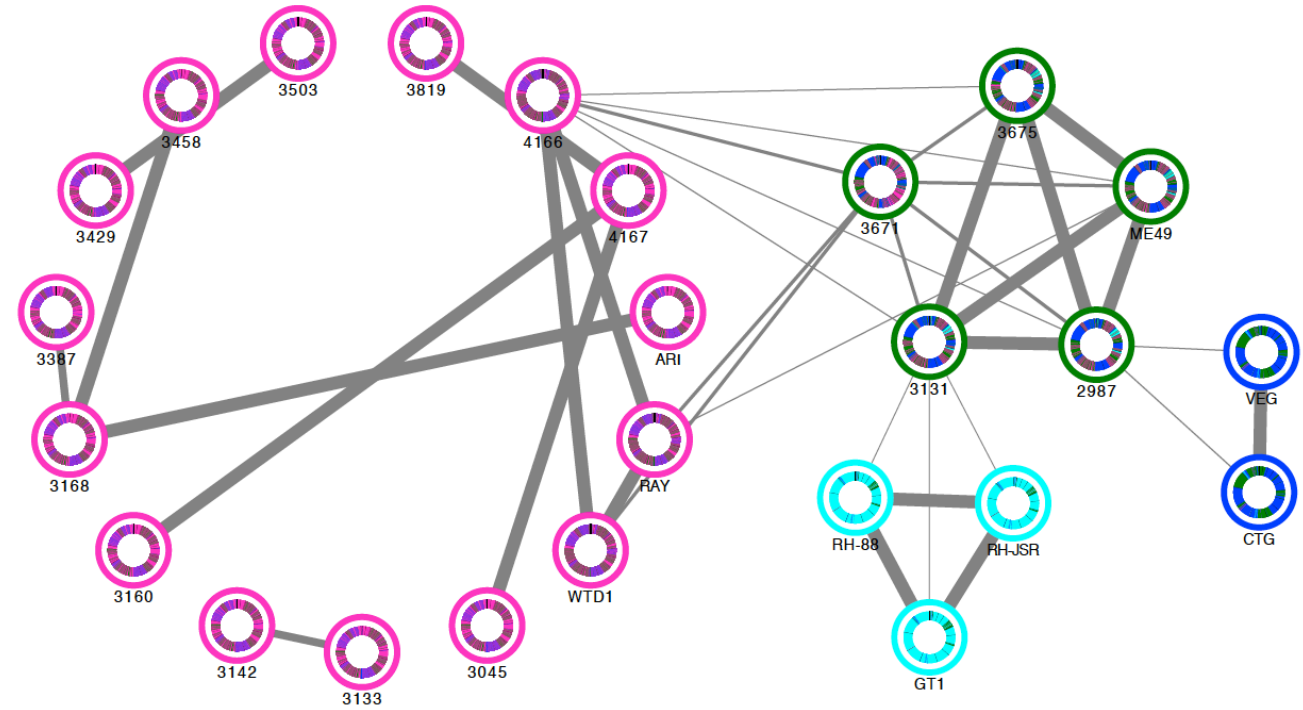
While the NeighborNet analysis established that Type X exists as a recombinant clade of strains, it failed to predict the precise number of genetic ancestries, or how they had admixed positionally across the chromosomes. PopNet was used to paint chromosomes according to their

local inheritance patterns (Zhang et al. 2017). Included in the analysis were all sequenced Type X and II strains, as well as reference Type I and III strains, known to have admixed with Type II. PopNet identified 4 distinct ancestries and each genome was a mosaic of the different ancestries, supporting the admixture model.

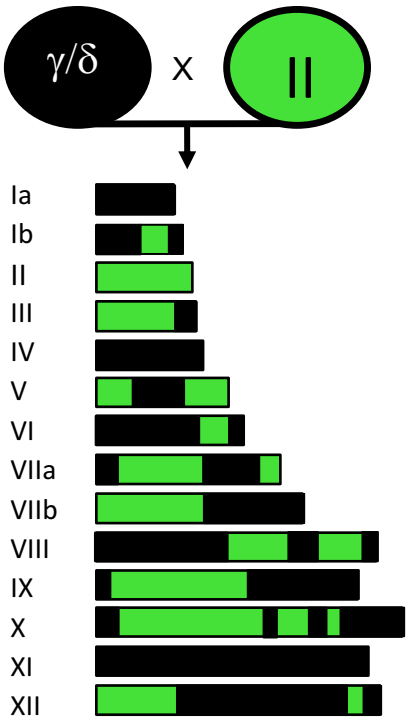
Within the circle depicting the Type X clade, 5 distinct subgroupings were identified based on the number and position of shared ancestral blocks; these were grouped together based on line thickness (Figure 5A). These same groupings were supported both by pairwise SNP plots and NeighborNet analysis, but the PopNet analysis showed which of the 5 ancestries had introgressed positionally across the mosaic genomes. Clear recombination blocks were readily resolved, and the genome architecture was remarkably similar between the groupings. A custom script designed to identify only major crossover points between the two parents (Type II and the mosaic ancestry of the γ/δ parent) identified either a limited number of single and double recombination events, or the inheritance of whole chromosomes of either Type II (Chr II) or γ/δ (Chr Ia, IV, XI) parental ancestry, consistent with this group of strains representing related sister progeny (Figure 5B). Because murine virulence plotted on the NeighborNet tree identified clear pathogenicity differences in mice that clustered based on parasite genotype, and because the recombinant clade of Type X strains resembled F1 progeny with only a limited number of crossover points, a forward genetic screen was attempted to determine whether genes contributing to the virulence trait could be mapped from the set of 18 natural isolates for which mouse virulence data existed (Figure 5C).

Figure 3-5: Whole genome sequencing demonstrates Type X is a unique clade of recombinant strains created from a cross between Type II parent and a unique γ/δ parent.
A) PopNet analysis of the WGS strains from A. The genome of each strain is represented by a circle of concatenated chromosomes which are painted in haploblocks based on their shared ancestry. All strains were clustered into four distinct color groups based on degree of shared ancestry (Type II-green, Type X-purple, Type I-cyan, Type III-blue). Line thickness between strains indicates their interrelatedness, bolder lines indicate whole genome linkage disequilibrium, as shown between the 4167 and 3160 strains. **B)** Genetic model of Type X parental recombination of isolate 3142. Type II ancestry is shown in green and unknown unique γ/δ parent in black. **C)** Isolates are shown with their group identification via 5 loci MLST (Figure 1A), eBURST clonal complex identity (Figure 1D), 17 loci MLST (Table 1), and murine virulence (Figure 1C).

A



B



C

Isolate	5 loci	CC	17 loci	Mouse VIR
2987	II	CC1	II	-
3131	II	CC1	II	-
3142	A	CC2	X1	+
3265	A	CC2	X1	+
3947	A	CC2	X1	n.d.
3168	A	CC3	X2	+
3458	B	CC3	X3	+/-
3523	B	CC3	X3	-
3026	C	CC4	X4	-
3045	C	CC4	X4	-
4167	C	CC4	X4	+/-
3160	C	CC4	X5	n.d.
3133	D	CC3	X6	+
3178	D	CC3	X6	n.d.
4166	D	-	X7	+
3819	E	-	X8	-
3503	F	CC2	X9	+/-
3429	F	CC2	X9	+/-
3387	F	CC3	X10	+/-
3675	G	CC1	X11	-
3671	H	-	X12	+

Natural Population-based QTL Identifies ROP33 as a Novel Murine Virulence Locus

The Type X isolates recovered from sea otters resembled a natural clade of recombinant F_1 progeny from a cross between a Type II strain, which is avirulent in mice, and an unknown parent that is a mosaic of two distinct ancestries. Among the natural isolates, a range of murine virulence was observed (Figure 1C). We performed a genome scan to determine the log-likelihood of association of discrete genome haploblocks with the acute virulence phenotype and identified four quantitative trait loci (QTL) peaks with logarithm of odds (LOD) scores 4.0 or greater on chromosomes V, VIIa, VIII, and X (Figure 6A). The average size of the genomic regions spanned by the QTLs were in the range of 100-200kb, except for the one on chromosome V, which was >700 kb (Figure 6B). To identify candidate genes within the four peaks, the following inclusion criteria were assessed: presence of a signal peptide and/or transmembrane domain, gene expression, polymorphism, genome-wide CRISPR score for essentiality (Sidik et al. 2016) (Supplemental Table 2). *ROP33* on chromosome VIIa stood out as the best candidate gene to target for reverse genetics as it was abundantly expressed during acute infection, it was predicted to be a functional serine-threonine protein kinase, and it was highly polymorphic. Of the shortlisted genes, one allele of *ROP33* was strongly correlated with acute virulence (p=0.00031; fishers two-sided exact test) (Figure 6C).

To determine if *ROP33* is a murine virulence gene, the *ROP33* locus was disrupted using targeted deletion by CRISPR-Cas9 facilitated double crossover homologous recombination in the mouse virulent RHΔ*ku80*Δ*rop18*Δ*hxgprt* strain. This strain is virulent in mice (LD_{100}=500 tachyzoites) and was engineered to accept targeted replacement of the *rop33* gene using an HXGPRT gene flanked by 30bp of homology just outside of the *rop33* promoter and 3'UTR region. Following selection in mycophenolic acid (MPA) and xanthine (to select for the

HXGPRT gene), the population was screened for disruption of the *rop33* gene by PCR. Mice were next infected with either the parent RHΔ*ku80*Δ*rop18*Δ*hxgprt* strain or the Rop33 mutant to assess murine virulence. Groups of 5 outbred CD-1 mice were injected intraperitoneally with 500 tachyzoites and the results shown are for one of two independent experiments. All mice infected with the parental RH strain succumbed to infection within 20 days (Figure 6D). Whereas, targeted deletion of the *rop33* gene was protective, indicating that ROP33 is an acute virulence gene. No difference in parasite load was detected through acute infection by bioluminescence imaging, indicating that deletion of ROP33 does not affect parasite proliferation *in vivo* (data not shown). All surviving mice were confirmed to be positive by serology, indicating that they had been productively infected. These data strongly suggest that ROP33 is a potent virulence factor for murine infection that was identified using recombinant progeny from a naturally occurring population genetic cross.

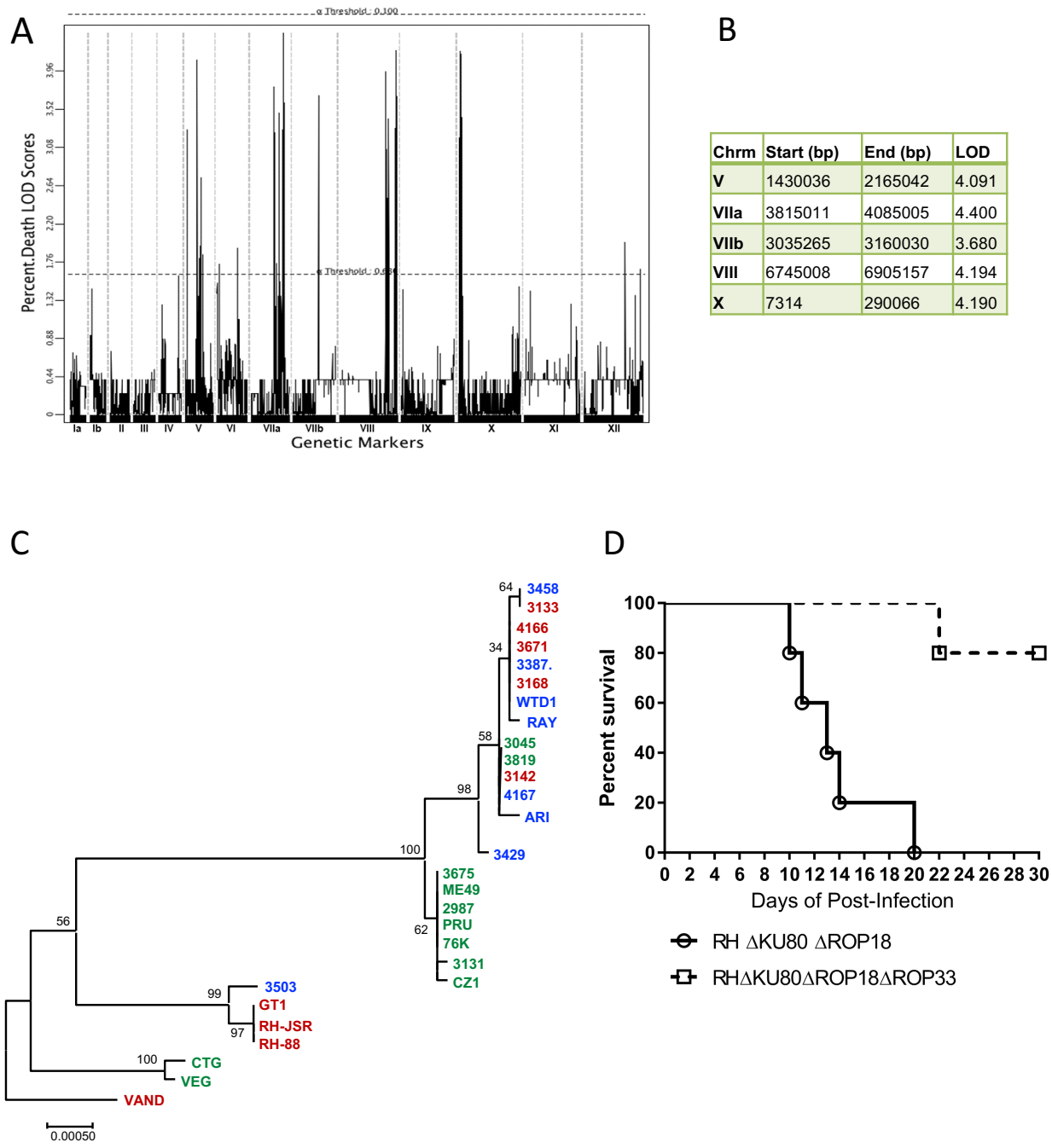


Figure 3-6: QTL analysis identifies ROP33 associated with low-dose murine virulence in natural population of Type X infections

A) WGS data from Type X strains was down-selected to one SNP every 5 kbp and analyzed via standard QTL methods with 1000 bootstrap support. LOD scores are shown across the chromosomes of these strains based on genetic association with low-dose murine virulence as shown in Table 1. **B)** Significantly associated peaks from the QTL were identified based on clusters SNP peaks as seen in the QTL analysis in A. The LOD scores of the tallest portion of the associated genomic region are listed. **C)** Maximum likelihood phylogenetic tree of ROP33 alleles from Type X strains displays that ROP33 allele correlates well with murine virulence. Colors correspond to murine virulence as listed in Figure 1C: virulent-red, intermediate virulent-blue, avirulent-green. **D)** ROP33 KO attenuates virulence in virulent Type I RH independently of ROP18.

Discussion

Here, we investigated a presumed epidemic outbreak of a clonal lineage (Type X) infecting sea otters to determine the genetic basis for this epizootic causing mortality in a federally listed, threatened species. Although Type X had previously been identified as the fourth clonal lineage in North America, work done here has shown that Type X is not a clonal lineage but rather a recombinant clade of strains that resemble F_1 progeny from at least one sexual cross. DNA sequence analysis of a wide selection of linked and unlinked markers across the nuclear and organellar genomes of Type X isolates identified Type X to be a composite of Type II and an unknown mosaic of two distinct ancestries, referred to as γ and δ, that have recombined to produce a highly invasive clade of strains causing infection and mortality in threatened marine mammals, including the Southern sea otter. Data shown here suggest that Type X was derived from one or a limited number of crosses from two sets of highly related parental strains. This suggests that *Toxoplasma*'s sexual life cycle is facilitating the evolution of new strains capable of expanding into new ecological niches, such as near-shore marine mammals off the Eastern Pacific coastline.

In previous studies, unknown strains having a distinct ancestry, referred to as α and β, were found to have crossed with Type II to create the Type I and Type III clonal clades respectively (Boyle et al. 2006). The pattern of recombined blocks inherited across the Type X strains is parsimonious with previously described models for the genetic history of Types I and III in which an unidentified ancestor (respectively α and β, and here designated γ and δ) sexually recombined with an ancestral Type II to create a new clade of strains (Figure 5B) (Fux et al. 2007; Boyle et al. 2006). Additionally, it has previously been suggested that Type X is a product of genetic hybridization between a Type II lineage and an unknown γ lineage, which has been

further refined in this analysis to be a mosaic of two distinct γ and δ ancestries (Miller et al. 2004; Sundar et al. 2008). The γ/δ lineage has likely recombined at least once with Type II to create the Type X recombinant clade of strains. Currently, no single isolate has been found which could be defined as either γ or δ, similarly to the α and β lines thought to have admixed with Type II to produce the clonal lines I and III (Boyle et al. 2006; Fux et al. 2007).

Previous work on Type X in wildlife samples using only a limited set of markers, identified Type X as the fourth clonal lineage in North America (Dubey et al. 2011; Khan, Dubey, et al. 2011). However, this designation was based on novel alleles identified at GRA6 and GRA7, which was the original basis for defining the Type X lineage as a new genotype that shared ancestry with Type II but had undergone a hybridization event with a novel genotype (Miller et al. 2004). It is now well established that Type X commonly infects wildlife across North America (Sundar et al. 2008; Dubey et al. 2011; Gibson et al. 2011; Su et al. 2012; VanWormer et al. 2014; Howe and Sibley 1995). This study examined *Toxoplasma* strains collected longitudinally across a seven-year period solely from sea otter infections and was comprised of infections that ranged in disease presentation between either asymptomatic to the primary cause of death. Previous work had classified Type X as clonal, so it was unexpected to find a myriad of different disease states in the infected sea otters. This investigation was carried out to determine the genetic basis for sea otter pathogenicity and to control for strain genotype as a critical parameter in the severity of disease (Dubey et al. 2011; Khan, Dubey, et al. 2011; Su et al. 2012; Sundar et al. 2008). Genotyping studies were carried out using an expanded set of markers to confirm the predicted clonal population structure for Type X. All genotyping markers used in this study were sequenced rather than examined strictly by PCR-RFLP analysis, as has been the standard for many previous studies. While the markers used herein only surveyed

a small part of the genome, they identified at least 12 distinct haplotypes within Type X, confirming that Type X is not a clonal population. Furthermore, at any given locus, only a Type II allele or one of two distinct alleles were identified indicating limited allelic diversity. Each of the 12 haplotypes possessed one of three allelic types at any one locus that had independently segregated across the markers examined. This result is parsimonious with the relationship between isolates as that of a recombinant clade of distinct ancestry. Our results were further confirmed both by CGH array of several genomes as well as whole genome sequencing and characterization. The misclassification as a clonal population appears to be the result of the low-resolution genotyping analyses performed, and the placement of the markers in predominantly Type II regions of the genome (Ajzenberg et al. 2002; Dubey et al. 2011; Howe and Sibley 1995). For this reason, future studies should use WGS to not only discriminate between haplogroups within the population, but also to determine the true clonality of each haplogroup. *Toxoplasma* population genetics would also greatly benefit from an increase in the number of isolates that are whole genome sequenced. WGS should be expanded to include both greater diversity of strains and a more in-depth analysis of the canonical clonal lineages.

The Type I and III clonal lineages are thought to be derived from a limited number of sexual crosses, and all strains within these clonotypes share large haploblocks of Type II-like sequence (Boyle et al. 2006; Grigg and Sundar 2009; Minot et al. 2012; Khan et al. 2014). This finding also holds true for the Type X lineage. At all loci surveyed, Type X displayed at least one isolate with a Type II inherited allele, and the introgression of Type II haploblocks across the genomes was confirmed in the 19 strains that underwent WGS. In regions that did not clade with Type II, two different allelic types were identified. We concluded that these represented distinct ancestries that we referred to as γ and δ. This designation was supported by the inherent

reticulation of the clade, seen via the NeighborNet tree (Figure 3B) which possessed short branch lengths from the reticulated network with multiple strains on a single branch, rather than producing a star phylogeny with multiple alleles present at each major branch, the result of strain-specific SNPs that accumulated through genetic drift. In fact, the lack of genetic drift between these loci supported genetic hybridization as the explanation for the relationship between the Type X haplotypes.

Sexual replication with genetic recombination is known to occur in South America at high frequency (Fux et al. 2007; Minot et al. 2012; Rajendran, Su, and Dubey 2012; Shwab et al. 2014). Additionally, recent studies on the population genetic structure of *Toxoplasma* has established extensive genetic hybridization at whole genome resolution for the vast majority of strains that represent the breadth of genetic diversity within *Toxoplasma* (Lorenzi et al. 2016; Zhang et al. 2017). Evidence of sexual recombination across the population as well as the evidence of sexual replication within the Type X clade shown here may indicate that sexual recombination is occurring with greater frequency than previously suspected within North America.

Sexual recombination is known to readily occur in the laboratory, although there is no current understanding of the genetic factors regulating sexual replication in the felid host (Behnke, Khan, and Sibley 2015; Grigg, Bonnefoy, et al. 2001; Saeij et al. 2006; Saeij et al. 2007; Taylor et al. 2006). It is unknown why sexual recombination does not appear to be occurring in North American wildlife as it is in South America, although it is possible that self-mating within the population masks the detection of genetic outcrossing (Ferguson 2002; Grigg and Sundar 2009; Wendte, Miller, Lambourn, et al. 2010). Especially with closely related

isolates, unisexual mating would not be resolved by current analyses and would be mistaken as asexual expansion (Wendte, Miller, Lambourn, et al. 2010; Wendte, Miller, Nandra, et al. 2010).

In order to expand its biological potential, *Toxoplasma* has been shown in the laboratory to utilize it sexual replicative life cycle to combine virulence alleles into new combinations that create enhanced virulence potential (Grigg, Bonnefoy, et al. 2001; Saeij et al. 2006; Taylor et al. 2006). Since sexual recombination can reshuffle alleles into new combinations that can produce a range of altered biological potential, the intermediate host range is likely responsible for selecting *Toxoplasma* strain diversity within each intermediate host (Grigg, Bonnefoy, et al. 2001; Saeij et al. 2006; Saeij et al. 2007; Taylor et al. 2006). Examples of this type of selection can be seen in murine hosts that have highly diversified IRG (immunity related GTPases) gene arrays. IRGs are the proteins primarily responsible for combatting *Toxoplasma* lysis of murine cells and *Toxoplasma* encodes a suite of highly polymorphic rhoptry kinase genes (ROPKs) whose primary function is to inactivate host IRGs (Fleckenstein et al. 2012; Hunn et al. 2011; Yamamoto et al. 2012). While laboratory mice are highly clonal, wild mice and their subsequent IRGs are more diverse, as are the ROPKs expressed by different *Toxoplasma* strains that are capable of infecting wild mice (Gazzinelli et al. 2014; Lilue et al. 2013). Similarly, TLR 11/12 is one of the primary rodent innate immune responses to detect infection by *Toxoplasma*. However, not all intermediate hosts share a functional combination of TLR11/12, thus, *Toxoplasma* strains that are selected for infection in mice need to bypass TLR11/12 recognition, whereas this is not a barrier to infection in a human cell, for example (Gazzinelli et al. 2014; Koblansky et al. 2013). Hence, reshuffling genetic diversity is one way in which the generalist parasite *Toxoplasma* finds the balance of fitness and infectious ability across its wide host range.

In sea otters, we have observed a host-virulence shift. The recombinant clade of Type X strains displays a wide spectrum of virulence within the sea otter host. One Type X subgroup (X9) causes severe myoencephalitis, and all otters infected with this sub-type die acutely. In contrast, another Type X subgroup (X1), that expanded in sea otters to represent greater than 50% of otter infections across a space of seven years, established benign, sub-clinical disease. While the X1 strain is relatively avirulent in sea otters, it was however highly pathogenic to mice, with all mice dying in 10-15 days with an inoculum of just 50 tachyzoites. This data supports an admixture model for the expansion and propagation of *Toxoplasma* in nature, whereby *Toxoplasma* utilizes its sexual cycle to create a wide spectrum of novel genotypes that possess altered biological potential that can be selected across *Toxoplasma*'s vast host range for the best parasite-host symbiosis. Thus, natural selection among intermediate hosts allows *Toxoplasma* to maintain strains that possess a cryptic virulence potential for another host species. This cryptic virulence is necessary in order to maintain its ability to cause disease outbreaks or expand its host range and transmissibility into new ecological niches.

In other protozoan parasites, when host partitioning occurs, it is synonymous with speciation. For example, *Plasmodium* parasites often partition by the mosquito host species they co-evolve with for parasite transmission (Mackinnon and Read 2004; Ariey et al. 2014; Bopp et al. 2013; Miotto et al. 2013; Arisue and Hashimoto 2014). Likewise, the apicomplexan parasite *Sarcocystis* largely maintains separate species for each intermediate-definitive host species combination to maintain its life cycle (Wendte, Miller, Lambourn, et al. 2010; Wendte, Miller, Nandra, et al. 2010; Barbosa et al. 2015). In contrast, *Toxoplasma* forms transmissible cysts in virtually all warm-blooded vertebrate hosts, and these are transmissible to the definitive host, as well as a large array of intermediate hosts. What is different, is that across the genetic diversity

of *Toxoplasma* strains, specific genotypes get selectively expanded in different intermediate hosts, and it is this fact that allows *Toxoplasma* to maintain cryptically virulent strains across its broad intermediate host range. Hence, the intermediate host is playing a central role in the natural selection, expansion, and maintenance of virulent strains across the broad host range of this generalist parasite. In effect, infected hosts act as reservoirs for the pathogenic or epidemic potential of the species. This study demonstrated that the sub-type of *Toxoplasma* that expanded to cause the majority of benign, chronic infections in sea otters was uniformly lethal to outbred laboratory mice. This is not unlike the expansion of Type I strains in birds that harbor asymptomatic infections of another highly pathogenic strain to laboratory mice (Miller et al, unpublished). Additionally, while Type II isolates are commonly found in domestic livestock in North America, Type X is more common in sylvatic hosts, leading others to propose that separate cycles exist within the *Toxoplasma* population that overlap solely in the feline definitive host (VanWormer et al. 2014; Wendte, Gibson, and Grigg 2011; Dubey et al. 2011).

Finally, although the exact parents are unknown that crossed to produce the recombinant clade of Type X strains, only a limited number of crossovers was detected, and the pedigree of one of the parents was known which allowed analysis by QTL on an unmanipulated, natural population of genetic hybrids. This analysis identified multiple punctate QTL associated regions with differences in mouse virulence. Taking advantage of the high-resolution SNP map for all of the progeny, and the ability to score quantitative differences in murine virulence, the regions of the genome that were associated with the virulence trait were small, allowing us to identify only a limited number of candidate loci for reverse genetic follow-up. The serine-threonine protein kinase ROP33 stood out as the best candidate to influence pathogenicity, largely because other related ROP proteins, including ROP5, 16, 17, and 18 are known to hijack host immune

signaling pathway gene expression to alter *Toxoplasma* virulence during rodent infection. Of note, all Type X isolates investigated in this study, whether virulent or avirulent, only expressed avirulent allele combinations of the known ROP virulence factors (data not shown). ROP33 was identified to be an active protein kinase that is abundantly expressed and is highly polymorphic. Whether its sequence divergence is the result of immune selection, or functional selection for interaction with specific host proteins is unknown. Nor is it known what specific host immune signaling pathways ROP33 targets. Although ROP33 contributes to the acute virulence phenotype, previous genetic mapping studies failed to identify the QTL for this locus, even though it is polymorphic between the Type I, II and III strains for which genetic crosses have been performed. It is likely that the ROP5 and ROP18 virulence factors were dominant, and ROP33 is analogous to ROP17, another virulence factor that was only identified after the virulence enhancing capacity of ROP18 was removed (Etheridge et al. 2014; Zhang et al. 2014). Future work will dissect the host factors that ROP33 targets to influence parasite transmission and pathogenicity.

Materials and Methods

Parasite Culture of Toxoplasma Strains

Fifty-three *T. gondii* strains were isolated from brain samples of stranded Pacific coast southern sea otters provided by Dr. Patricia Conrad as previously described (Miller et al. 2004). Cause of death was determined for each sea otter host from primary pathology reports done based on host necropsy (Kreuder et al. 2003; Miller et al. 2002). *Toxoplasma* isolates are

identified based on their ATOS ("as the otter swims") numbers derived from sea otter host stranding location as previously described (Wendte, Miller, Nandra, et al. 2010).

Parasite strains were cultured on human foreskin fibroblast (HFFs) monolayers at 5% CO_2 at 37°C. These HFFs were grown in DMEM Complete Medium derived from Dulbecco's Modified Eagle Medium (DMEM) enriched with 10% heat-inactivated fetal bovine serum (FBS), glutamine, and treated with gentamycin and penicillin-streptomycin as described previously (Pszenny et al. 2000). *Toxoplasma* strain tachyzoites were grown on host HFFs until monolayer lysis.

Type X Mouse Virulence Assay

To determine the virulence of these isolates in a mouse model, Type X strain virulence was characterized using low dose infection in outbred mice. Groups of five or more, 6-8 week-old, female CD-1 mice from Charles River Laboratories were intraperitoneally injected with 50 *T. gondii* tachyzoites per strain suspended in 500 µl PBS (Su et al. 2002). Mice were weighed daily to measure infection-induced cachexia and mouse survival was assayed over 42 days. After 14 days, mice were bled, and serum extracted to test for seroconversion via indirect fluorescent antibody test (IFAT) against ME49 tachyzoites (Fletcher 1965; Grigg, Bonnefoy, et al. 2001; Miller et al. 2002). Virulence in murine infection was defined as thus: Virulent strains killed all seropositive mice, avirulent strains killed none of the seropositive mice, and intermediate virulence strains killed around 50% of all seropositive mice infected per strain of Type X.

DNA Extraction and Type X Strain Preliminary Marker Typing

Parasites were syringe lysed from host HFFs using a 27-gauge needle and filtered through a 3.0-micron polycarbonate filter to remove cellular debris. Cleaned tachyzoites were pelleted and rinsed with PBS. DNA was extracted from cell pellets using the Qiagen DNeasy Blood and Tissue kit (Qiagen).

DNA from the 53 filtered isolates was typed using 5 previously described PCR-RFLP markers distributed across 5 chromosomes (Su, Zhang, and Dubey 2006; Grigg, Bonnefoy, et al. 2001; Fazaeli et al. 2000; Howe and Sibley 1994; Miller et al. 2004). BSR4, BAG1, GRA6, ROP1, and SAG3 were amplified by PCR using Taqman Ampli-Taq polymerase and an Eppendorf thermocycler as previously described (Fazaeli et al. 2000; Su, Zhang, and Dubey 2006; Sundar et al. 2008). PCR products were electrophoresed through a 1% agarose gel to verify size. The PCR product of ROP1 was restriction enzyme digested with DdeI for 1 hour at 37°C, products were run on a 1% agarose gel, and compared to clonally representative strain digests to determine marker identity. The PCR products of BSR4, BAG1, GRA6, and SAG3 were Sanger sequenced and sequences were analyzed using DNAStar's Lasergene SeqMan Pro software version 14 to identify representative SNPs within these genes, detailed below. These five markers identified 10 unique genotypes within the Type X clade.

21 strains representing these 10 unique genotypes were PCR amplified and Sanger sequenced at 20 genotyping markers across the parasite genomes: 18 markers were sequenced for all possible strains at both linked and unlinked genomic loci, encompassing 13 of the 14 chromosomes, and 2 loci on organellar genomes (apicoplast and mitochondria) (see Supplemental Table 1). Only seven loci on different Type X strains (shown in gray) were unable

to be amplified and characterized. 15,430 bp of genome were sequenced from these makers with 335 SNPs (~0.024% of the genome). Sequences were examined, and nucleotides verified using SeqMan Pro alignment software (Lasergene). Sequences for reference strains (ME49, GT1, and VEG) were downloaded from ToxoDB Version 8.2 (Gajria et al. 2008). Alleles at each locus were determined by phylogenetic tree analysis as detailed below.

eBURST Clonal Complex Determination

Alleles determined from the 17 nuclear encoded marker sequencing (excluding ROP1 due to its classification as a microsatellite marker) were categorized based on phylogenetic analysis of the sequence at a given locus and given numerical designations to create a multilocus sequence-typing scheme suitable for eBURST (Feil et al. 2004). Default settings were used to evaluate the isolated strains' alleles to detect clonal complexes within the population of strains. Strains sharing 16 of the 17 nuclear encoded markers were designated as clonal complexes and represented by connecting lines in the diagram shown.

Phylogenetic Tree Creation and Marker Analysis

Typing marker sequences from Sanger sequencing were aligned using Clustal X 2.1 with default settings (Larkin et al. 2007). For each marker, the alignment was input into MEGA 7 to create a maximum likelihood tree using Tamura-Nei model distance analysis with uniform rates of substitution across all sites, and 1000 bootstrap support for all branch points (Tamura et al. 2013). Bootstrap support equal to or greater than 60% distinguished novel alleles from

mitotically drifted alleles. This cutoff value was determined based on previously characterized markers such as GRA6 that were initially used to separate the Type X clade from the closely related Type II clade using single marker phylogenetic trees (Khan, Dubey, et al. 2011). All typing marker phylogenetic trees were rooted the canonical Type II strain, ME49 due to its prediction to be ancestral to the Type X clade.

Distinct parental lineages with bootstrap support greater than 60% are indicated on each tree, where II indicates the ancestral Type II lineage and γ and δ indicating the yet unidentified secondary parent(s) of the Type X strains. These novel alleles are indicated in purple and orange respectively on the 17 loci MLST in Supplementary Table 2.

Genomic Hybridization to Affymetrix Arrays

Genomic DNA was sheared, biotin labeled, and then hybridized to a custom *T. gondii* Affymetrix microarray as previously published (Khan, Miller, et al. 2011). High-fidelity SNPs were characterized via a custom R script to identify SNPs belonging to each of the three reference strains (I, II, and III) which are color coded according to allelic similarity to the reference strains. Three reference strains (GT1, ME49, and CTG) are shown to demonstrate ideal hybridization within canonical lineages. Sea otter Type X DNA isolate hybridizations are shown under these.

Characterization of Type X Whole Genome Diversity

DNA from 16 isolates of Type X that grew best in culture and 3 representative Type X strains (ARI, RAY, and WTD1), for a total of 19 Type X strains, was collected and isolated as stated above. 1 ng of DNA from each of these 19 Type X strains was sent to Rocky Mountain Laboratories for whole genome sequencing using Illumina HiSeq technology. Fastq reads derived from Illumina sequencing were reference mapped against the ToxoDB version 8.2 ME49 genome using BWA 0.7.5a to align the reads to the reference genome and GATK 3.7 in coordination with Picard 1.131 following best practices to quality control the mapped reads (Van der Auwera et al. 2013). Following mapping, the gVCF method of GATK was used to combine SNP calls using stand_call_conf of 30.0, nct of 10, and ploidy of 1 (for haploid genomes) to call 568592 single nucleotide polymorphism positions across the whole genomes of these strains (Van der Auwera et al. 2013). The derived VCF formatted SNP file was curated using GATK and VCFTools to produce a tabular file containing only biallelic SNPs with no large insertions or deletions (Danecek et al. 2011).

A custom script was utilized to convert this SNP file into a fasta file of strain polymorphic positions across the Type X and reference genomes. These SNP fastas were input into SplitsTree4 where the default parameters for BioNJ with 1000 bootstrap support was used to create a NeighborNet tree (Huson and Bryant 2006). Interconnected reticulation between strains is indicative of recombination between strains while mitotic drift can be seen by divergent branching.

SNP Density Fingerprint Analysis of Recombinant Progeny

The same tabular, biallelic SNP file used in the NeighborNet analysis was modified using a custom R script based on location mapping derived from Behnke et al. 2015 to isolate strain-specific polymorphic locations from the tabular VCF. Further R scripts group these SNPs into 100 kbp windows which are mapped across the genomes of these strains to display where diversity from the reference genome (ME49) is most apparent (Minot et al. 2012; Yin, Cook, and Lawrence 2012). The larger the number of SNPs in a particular window, the more divergent this strain is from the ME49 reference. Distinct haploblocks where genomic recombination has occurred are apparent in areas where haploblock diversity significantly varies from the surrounding regions on the same chromosome.

PopNet Characterization of Strain Interrelatedness

The tabular, biallelic SNP file created from WGS of Type X and used to derive the SNP density plots and NeighborNet tree was input into PopNet using default parameters to assess the diversity and interrelatedness of strains (Zhang et al. 2017). The recombination and Markov clustering output were input into Cytoscape for visualization. Genomes are displayed in circularized format with chromosomes concatenated into a circular genome display. The background of each strain is painted to match the group that shares the most common ancestry over the entire genome with the strain shown. Chromosome painting is done on strain genomes such that haploblocks of 10 kbp of sequence are painted based on shared ancestry within the haploblock. For instance, a Type X genomic haploblock that is closely related to the ancestral

Type II haploblock at will be painted green to indicate inheritance. Strains that are more closely related have thicker connecting lines between the circles of the strain isolations.

QTL Virulence Candidate Gene Identification

To find novel virulence alleles, the tabular, biallelic SNP file used previously is utilized to run quantitative trait loci (QTL) analysis on these strains of Type X *Toxoplasma* (Behnke et al. 2015). SNP calls were down-selected by a custom java script to include one SNP every 5 kb. This allows the QTL software to analyze the breadth of the WGS data in QTL, which was built to handle marker typing data without the depth inherent in WGS data. Custom scripts code the SNPs into reference (ME49) versus alternative (Type X's γ/δ lineage which will substitute as the secondary parent) alleles. This down-selected marker dataset was combined with previously determine low-dose murine virulence for these strains. This data was then input into J/qtl and a one QTL genome scan was run using the default settings for the EM algorithm (maximum likelihood) with 1000 permutations to identify genomic locations significantly associated with the percentage of murine death with Type X infection (Smith et al. 2009).

From the QTL calculations, four significantly associated regions were identified based on a LOD score of greater than 4.0. Within these genomic regions, 450 genes were predicted based on ToxoDB documentation. These potential genes were down-selected based on presence of a signal peptide and/or transmembrane domain, gene expression, polymorphism, and genome-wide CRISPR score for essentiality to identify 32 virulence candidate genes within the Type X strains. Of these, ROP33 was selected for further interrogation based on its high LOD score and similarity to previously identified virulence effector proteins.

ROP33 Knock-out and Virulence Assay

WGS fastqs of the Type X strains and several reference strains were reference mapped to the ME49 allele of ROP33 downloaded from ToxoDB. SNPs within these strains were characterized as with WGS above, and using GATK and BEDTools, consensus sequences were called to create fastas of this virulence candidate gene in each strain (Quinlan and Hall 2010; Van der Auwera et al. 2013). These fasta files were compared to one another using SeqMan Pro 14.0 and polymorphisms were used to characterize the allelic inheritance of these strains by creating phylogenetic trees as with the Sanger sequenced markers above.

ROP33's effect on virulence was determined by knocking out the ROP33 gene in a ROP18 deficient Type I strain of *Toxoplasma.* ROP33 was deleted from the virulent RHΔ*ku80*Δ*rop18*Δ*hxgprt* strain by CRISPR-Cas9 targeting using an HXGPRT gene flanked by 30bp of homology just outside of the *rop33* promoter and 3'UTR region. The resultant RHΔ*ku80*Δ*rop18*Δ*hxgprt*Δ*rop33* population screened by PCR for deletion and was selected using mycophenolic acid (MPA) and xanthine followed by cloning to produce a clonal deletion of ROP33.

To assess the virulence of ROP33, groups of 5 outbred CD-1 female 6-8 week old mice were injected intraperitoneally with 500 tachyzoites were infected with either the parent RHΔ*ku80*Δ*rop18*Δ*hxgprt* strain or the Rop33 mutant. Infectivity and virulence were assessed as above. Parasite dissemination and growth were assessed by bioluminescence imaging as previously described (Saeij et al. 2005).

Ethics Statement

The use of animals in this study was done under the approval of the Animal Care and Use Committee of the National Institute of Allergy and Infectious Diseases, National Institutes of Health (LPD-22E). All guidelines required by the NIH and the Animal Welfare Act were strictly followed.

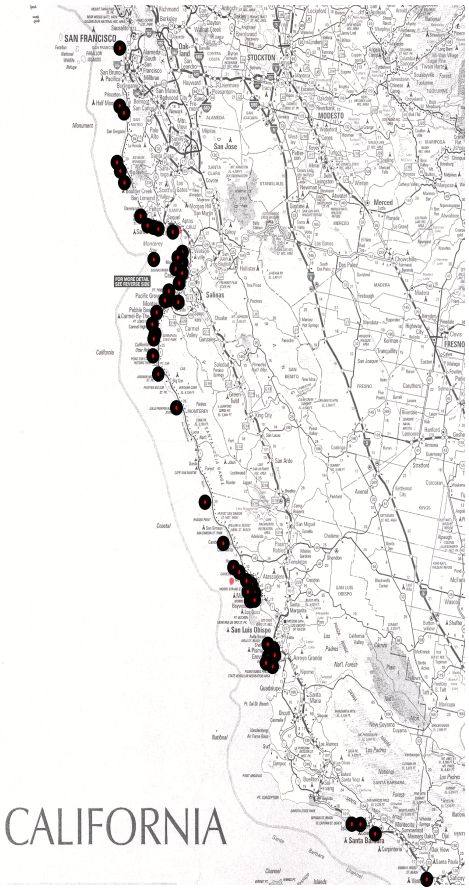
Acknowledgements

Andrea Packham and Ann Melli for their assistance in culturing and isolation of the *Toxoplasma* isolates. Natarajan Sundar for isolation and preliminary characterization of *Toxoplasma* strains.

Supplemental Figures

Supplemental Figure 3-1: Stranding location and probably cause of death of sea otter infections.
A) Geographical map of all 53 Toxoplasma-infected sea otter stranding locations along the California coast line. Each sea otter is represented by a red dot located at stranding location. **B)** Isolates of Toxoplasma gondii used in this paper. Causes of death (COD) are listed based on pathology determination at date of necropsy. ATOS numbers and stranding locations identify the discovery location of the sea otter. Grouping from 5 and 17 markers are shown as in Figure 5C.

A



B

Isolate	ATOS #	5 loci	17 loci	Stranding Location	Necropsy Date	Sex	IFAT	Primary COD	Secondary COD	Tertiary COD	Infection
2994	249	II		Santa Cruz	05/15/98	M	Serum	Shark Attack - Presumptive	None	None	Tg
3739	264	II		Santa Cruz 26th Ave Beach	06/12/02	M	2560	Meningoencephalitis - Protozoal	Boat Strike	Emaciation/Starvation	Tg
3576	274	II		New Brighton State Beach	09/05/01	M	20480	Shark Attack - Presumptive	Meningoencephalitis - Idiopathic	Myocarditis	Tg
3587	281	II		Rio Del Mar Beach	09/22/01	M	1280	Shark Attack - Presumptive	Meningoencephalitis - Protozoal	Ulcers - Gastric	Tg
2987	291	II	II	Mannressa State Beach	05/11/98	F	1280	Shark Attack - Presumptive	Meningoencephalitis	Emaciation/Starvation	Tg
4181	303	II		Sunset State Beach	04/14/04	F	40960	Bacterial - Abcess	Pneumonia - Bacterial	Meningoencephalitis - Protozoal	Sn/Tg
3009	321	II		Elkhorn Slough	06/04/98	M	640	Boat Strike - Presumptive	Bacterial - Abcess	None	Tg
3521	325	II		Salinas River State Beach	05/19/01	M	2560	Hemorrhage - Brain	Meningoencephalitis - Protozoal	Cardiomyopathy - Idiopathic	Tg
3208	357	II		Monterey Bay	07/12/99	M	5120	Drowning - Presumptive	Meningoencephalitis - Idiopathic	None	Tg
3636	376	II		Del Monte Beach	12/27/01	M	80	Shark Attack - Presumptive	Ulcers - Gastric	Meningoencephalitis - Idiopathic	Tg
3087	377	II		Monterey Harbor	10/28/98	F	640	Acanthocephalan Peritonitis	Mating Trauma	Emaciation/Starvation	Tg
3131	379	II	II	Monterey Bay	02/05/99	M	320	Cardiomyopathy	Mating Trauma	Emaciation/Starvation	Tg
3005	928	II		Pismo Beach	06/01/98	M	160	Presumptive	Ulcers - Gastric	None	Tg
3396	928	II		Oceano Dunes	11/11/00	M	20480	Drowning - Presumptive	Meningoencephalitis - Idiopathic	None	Tg
3097	84	A		Dunes Beach, San Mateo	11/16/98	M	5120	Meningoencephalitis - Idiopathic	Acanthocephalan Peritonitis	Emaciation/Starvation	Tg
3142	180	A	X1	Ano Nuevo Island	02/24/99	F	320	Shark Attack - Presumptive	Meningoencephalitis - Protozoal	Ulcers - Gastric	Tg
3786	180	A		Ano Nuevo Island	09/20/02	M	2560	Shark Attack - Confirmed	Systemic Protozoal Infection	Ulcers - Gastric	Sn
4045	260	A		Seabright Beach	10/21/03	M	10240	Shark Attack - Presumptive	None	None	Tg
3950	263	A		Santa Cruz, 21st Street Beach	07/16/03	M	10240	Meningoencephalitis - Protozoal	Cardiomyopathy	Septicemia - Bacterial	Sn/Tg
3520	306	A		Sunset Beach	05/16/01	M	320	Acanthocephalan Peritonitis	Enteritis	Emaciation/Starvation	Tg
3659	374	A		Del Monte Beach	02/28/02	F	40	Intussusception	Peritonitis - Septic	Septicemia - Bacterial	Tg
3265	426	A	X1	Carmel Beach	09/28/99	F	10240	Acanthocephalan Peritonitis	Septicemia - Bacterial	Herniation	Tg
3168	450	A	X2	Monterey Bay	04/09/99	F	2560	Gastroenteropathy - Hemorrhagic	Emaciation/Starvation	Rhabdomyolysis/Necrososis	Tg
3865	480	A		Carmel River State Beach	04/03/03	F	10240	Intestinal Volvulus	Mating Trauma	Emaciation/Starvation	Tg
3947	523	A	X1	Pfieffer Burns Beach	06/26/03	F	20480	Nose Wound	Septicemia - Bacterial	Emaciation/Starvation	Tg
4003	750	A		Cambria	09/17/03	F	10240	Mating Trauma	Meningoencephalitis - Idiopathic	Emaciation/Starvation	Tg
4071	807	A		Cayucos, North of E Steet	12/03/03	F	2560	Nose Wound	Septicemia - Bacterial	Emaciation/Starvation	Tg
3483	820	A		Morro Strand	03/30/01	F	320	Hepatopathy	Cerebral Edema	Pneumonia - Bacterial	Sn/Tg
4151	825	A		Morro Bay	04/04/04	F	>40960	Meningoencephalitis - Protozoal	Myocarditis - Nonsuppurative	Emaciation/Starvation	Sn/Tg
3488	919	A		Pismo Beach, ~400m North of Grand Ave	04/04/01	F	640	Cardiomyopathy - Idiopathic	Emaciation/Starvation	Encephalitis - Atypical Idiopathic	Tg
3821	922	A		Pismo Beach	01/06/03	M	2560	Meningoencephalitis - Protozoal	Cardiomyopathy	Drowning - Presumptive	Tg
3744	924	A		Pismo Beach	07/02/02	F	5120	Shark Attack - Confirmed	None	None	Tg
3637	926	A		Oceana Dunes	02/14/03	F	No data	Meningoencephalitis - Protozoal	Myocarditis	Septicemia - Bacterial	Tg
3523	141	B	X3	Pescadero Point	05/31/01	F	5120	Meningoencephalitis - Protozoal	Acanthocephalan Enteritis	None	Sn/Tg
3458	422	B	X3	Stillwater Cove	01/16/01	F	<80	Meningoencephalitis - Protozoal	Cardiomyopathy	Emaciation/Starvation	Tg
3897	435	B		Pt Lobos State Reserve	04/22/03	F	10240	Mating Trauma	Pyothorax	Septicemia - Bacterial	Tg
3728	1,445	B		Ventura County (Pt Mugu St Pk)	05/29/02	M	320	Cardiac Dilation	Endocarditis - Vegetative	Domoic Acid Intoxication - Presumptive	Tg
3077	145	C		Bean Hollow, Half Moon Bay	09/27/98	M	1280	Shark Attack - Presumptive	None	None	Tg
3045	808	C	X4	Cayucas	07/09/98	F	Serum	Presumptive	Heart Failure - Chronic	Emaciation/Starvation	Tg
3026	840	C	X4	Morro Bay Sandspit	06/24/98	M	320	Domoic Acid Intoxication	Blunt Trauma	Coccidiomycosis	Tg
4167	852	C	X4	Morro Creek	04/09/04	M	10240	Meningoencephalitis - Protozoal	Cardiomyopathy	Bacterial - Abcess	Sn/Tg
3160	933	C	X5	Pismo Beach	03/29/99	M	320	Bacterial - Abcess	Cardiac Dilation	Meningitis - Idiopathic	Tg
3178	321	D	X6	Moss Landing Harbor	03/26/02	F	320	Advanced decomp-No histopath	Advanced decomp-No histopath	Advanced decomp-No histopath	Tg
3183	816	D		Morro Strand	05/04/99	F	640	Meningoencephalitis - Idiopathic	Meningoencephalitis - Protozoal	Acanthocephalan Peritonitis	Tg
3133	818	D	X6	Morro Strand Campground	02/04/99	M	20480	Emaciation/Starvation	Meningoencephalitis - Protozoal	Viral Infection	Tg
4166	827	D	X7	Morro Bay	04/09/04	M	40960	Meningoencephalitis - Protozoal	Acanthocephalan Peritonitis	Cardiomyopathy	Sn/Tg
3451	836	D		Morro Bay, Sandspit between port 3 & 4	11/29/00	M	<80	Meningoencephalitis - Protozoal	Cardiomyopathy	None	Tg
3819	732	E	X8	San Simeon Bay	12/28/02	F	5120	Pneumonia - Bacterial	Mating Trauma	Peritonitis - Septic	Tg
3429	827	F	X9	Morro Bay	10/30/00	F	2560	Acanthocephalan Peritonitis	Septicemia - Bacterial	Herniation	Tg
3387	922	F	X10	Pismo Beach	07/19/00	F	1280	Meningoencephalitis - Idiopathic	Meningoencephalitis - Protozoal	Acanthocephalan Peritonitis	Tg
3503	1,319	F	X9	Rincon Pt., Ventura County	08/09/01	M	2560	Meningoencephalitis - Protozoal	None	None	Tg
3675	435	G	X11	Point Lobos	03/20/02	F	1280	Emaciation/Starvation	Nose Wound	Meningoencephalitis - Protozoal	Tg
3671	321	H	X12	Elkhorn Slough	03/13/02	M	640	Pneumonia - Bacterial	Cardiomyopathy	Meningoencephalitis - Protozoal	Sn/Tg

Marker	5 Loci	17 Loci	C22B	Chr_II_10	C29Z	BSR4	L35B	PK1	SAG4	ROP18	GRA7	BAG1	SR52	SAG1	SAG2	BTB	GRA6	ROP1	UPRT	SAG3	APICO	COX1
Chromosome			Ib	II	III	IV	V	VI	VIIa	VIIa	VIIa	VIIb	VIII	VIII	VIII	IX	X	XI	XI	XII	API	MITO
Length			421	639	397	1138	354	1011	174	1617	460	1917	983	700	1243	331	758	960	1214	253	442	418
Type I		I																				
Type II		II																				
Type III		III																				
2987	II	II																				
3131	II	II																				
3142	A	X1																				
3265	A	X1																				
3947	A	X1																				
3168	A	X2																				
3458	B	X3																				
3523	B	X3																				
4167	C	X4																				
3026	C	X4																				
3045	C	X4																				
3160	C	X5																				
3133	D	X6																				
3178	D	X6																				
4166	D	X7																				
3819	E	X8																				
3503	F	X9																				
3429	F	X9																				
3387	F	X10																				
3675	G	X11																				
3671	H	X12																				

Type II	
	32
	44

Supplemental Table 3-1: Expanded sequencing marker genotyping of Type X isolates reveals chromosomal segregation and recombination within chromosomes.
21 isolates characterized at 17 nuclear markers (ROP1 was excluded as it is a microsatellite marker, and prone to elevated mutation rates) clade into 12 distinct Type X genotypes (X1-X12). Types I, II, III, γ and δ lineage alleles, as determined by phylogenetic comparisons, are colored red, green, blue, purple, and orange, respectively. Shades of colors represent genetic drift below the 60% bootstrap delineation from the canonical allele. White represents uninformative or incomplete DNA sequencing results.

Chr	Start (bp)	End (bp)	LOD	TM	Signal	Gene ID	Genomic Location(s)	Product Description
5	1215032	2165042	4.09	Y	Y	TGME49_285870	TGME49_chrV:2,158,544..2,162,541(-)	SAG-related sequence SRS20A
				Y	Y	TGME49_285940	TGME49_chrV:2,123,872..2,135,543(-)	hypothetical protein
				Y	Y	TGME49_286180	TGME49_chrV:1,933,873..1,942,288(-)	tRNA ligases class I (M) protein
				Y	Y	TGME49_286450	TGME49_chrV:1,824,587..1,826,598(+)	dense granule protein GRA5
				Y	Y	TGME49_286530	TGME49_chrV:1,790,919..1,793,716(+)	hypothetical protein
				Y	Y	TGME49_286620	TGME49_chrV:1,754,511..1,762,820(-)	S1 RNA binding domain-containing protein
				Y	Y	TGME49_286630	TGME49_chrV:1,750,888..1,752,970(+)	redoxin domain-containing protein
				Y	Y	TGME49_286770	TGME49_chrV:1,682,413..1,683,569(+)	hypothetical protein
7a	2615053	4085005	4.35	Y	Y	TGME49_201130	TGME49_chrVIIa:4,019,644..4,026,659(+)	rhoptry kinase family protein ROP33
				Y	Y	TGME49_201180	TGME49_chrVIIa:3,982,746..3,988,716(-)	hypothetical protein
				Y	Y	TGME49_201390	TGME49_chrVIIa:3,908,769..3,911,287(-)	hypothetical protein
				Y	Y	TGME49_201780	TGME49_chrVIIa:3,794,121..3,797,385(-)	microneme protein MIC2
				Y	Y	TGME49_201860	TGME49_chrVIIa:3,741,150..3,743,832(-)	hypothetical protein
				Y	Y	TGME49_202050	TGME49_chrVIIa:3,666,573..3,668,278(-)	hypothetical protein
				Y	Y	TGME49_202375	TGME49_chrVIIa:3,379,488..3,384,231(-)	hypothetical protein
				Y	Y	TGME49_202440	TGME49_chrVIIa:3,326,760..3,330,921(+)	hypothetical protein
				Y	Y	TGME49_202572	TGME49_chrVIIa:3,207,247..3,216,124(-)	ribophorin i protein
				Y	Y	TGME49_202620	TGME49_chrVIIa:3,164,535..3,167,304(-)	hypothetical protein
				Y	Y	TGME49_202630	TGME49_chrVIIa:3,149,594..3,162,966(-)	ATP-dependent metallopeptidase HflB subfamily protein
				Y	Y	TGME49_202650	TGME49_chrVIIa:3,120,640..3,134,651(+)	hypothetical protein
				Y	Y	TGME49_202780	TGME49_chrVIIa:3,033,153..3,037,960(-)	rhoptry kinase family protein ROP25
				Y	Y	TGME49_202800	TGME49_chrVIIa:3,017,549..3,023,257(-)	cytochrome c oxidase assembly protein COX11 protein, putative
				Y	Y	TGME49_202860	TGME49_chrVIIa:2,977,558..2,979,383(+)	hypothetical protein
				Y	Y	TGME49_202940	TGME49_chrVIIa:2,936,249..2,939,562(-)	hypothetical protein
8	6470103	6905157	3.94	Y	Y	TGME49_200230	TGME49_chrVIII:6,753,611..6,755,422(-)	microneme protein MIC17C
				Y	Y	TGME49_200440	TGME49_chrVIII:6,883,172..6,887,049(+)	hypothetical protein
10	7314	595006	3.95	Y	Y	TGME49_228110	TGME49_chrX:579,227..581,518(-)	hypothetical protein
				Y	Y	TGME49_228160	TGME49_chrX:528,230..538,369(-)	acid phosphatase
				Y	Y	TGME49_228170	TGME49_chrX:519,357..528,263(-)	inner membrane complex protein IMC2A
				Y	Y	TGME49_228310	TGME49_chrX:426,285..427,726(-)	hypothetical protein
				Y	Y	TGME49_228350	TGME49_chrX:381,741..397,668(+)	elongation factor Tu GTP binding domain-containing protein
				Y	Y	TGME49_228360	TGME49_chrX:376,764..379,771(+)	peptidyl-prolyl isomerase FKBP12, putative

Supplemental Table 3-2: Potential virulence candidate genes derived from the QTL analysis.
Chromosome peaks with LOD scores over 3 as shown in Figure 3B were interrogated via ToxoDB to select genes in these regions which have predicted signal sequences (Signal) and transmembrane (TM) domains in Toxoplasma. Genes are listed with gene ID, genomic location, and predicted protein function.

Chapter 4 - Low Frequency Sexual Recombination in *Toxoplasma* Favors Clonal Expansion by Unisexual Mating

Kennard, Andrea; Gregg, Beth; and Grigg, Michael E.

Laboratory of Parasitic Diseases, NIAID, National Institutes of Health, Bethesda, MD.

Introduction

Toxoplasma gondii is a generalist parasite found worldwide and is capable of infecting any warm-blooded host. This parasite has a bimodal lifecycle that allows *Toxoplasma* to replicate both asexually and sexually. The asexual cycle propagates *Toxoplasma* between intermediate hosts either vertically by congenital infection or horizontally by carnivory. Conversely, sexual propagation of *Toxoplasma* within its definitive feline host promotes either genetic diversification via outcrossing to produce novel admixture strains or sexual expansion of a single clone by self-mating. Both outcomes produce up to 100 million infectious oocysts per feline. Novel admixtures can then be naturally selected within the parasite's broad range of intermediate hosts to select for best-fit combinations between host and parasite that maximize transmissibility (Sibley et al. 2009). Given the fecundity of the *Toxoplasma* sexual cycle, the expectation was that its population genetic structure would be highly diverse. But this is not the case in North America and Europe which possesses a limited number of clonal haplogroups or clades that cause the majority of infections in people and livestock (Boyle et al. 2006; Minot et al. 2012; Howe and Sibley 1995; Khan, Dubey, et al. 2011; Shwab et al. 2014; Dubey et al. 2014; Dubey et al. 2011; Su, Zhang, and Dubey 2006; Grigg and Sundar 2009). It was previously hypothesized that these clonal lineages may be the products of a recent selective sweep or

founder effect, whereby only a few parasite genotypes escaped a genetic bottleneck and founded the current clonal population structure (Bertranpetit et al. 2016; Fux et al. 2007; Khan, Miller, et al. 2011; Shwab et al. 2014). Alternatively, a population structure with limited genomic diversity could also be explained by parasites that are predominantly expanded asexually and only rarely outcross in nature (Grigg and Sundar 2009; Khan et al. 2006; Sibley and Ajioka 2008). However, this hypothesis is at odds with the high fecundity with which *Toxoplasma* laboratory strains undergo experimental outcrossing in the feline host. Recent work by Wendte et al, 2010 established that *Toxoplasma* frequently undergoes either unisexual mating between two genetically similar strains from the same clade or self-mating, in which a single strain undergoes a full sexual cycle within the definitive host to produce clonal progeny (Wendte, Miller, Lambourn, et al. 2010). In this way, unisexual or self-mating allows successful clones of the parasite to expand their prevalence and host range within a geographical or ecological niche which serves to genetically homogenize diversity within a given population (Heitman 2010; Roach and Heitman 2014; Wendte, Miller, Lambourn, et al. 2010).

The designation that *Toxoplasma* possesses a clonal genetic population structure in North America and Europe was established using solely low-resolution typing markers (Su, Zhang, and Dubey 2006; Ajzenberg et al. 2010; Blackston et al. 2001; Verma et al. 2015). Subsequent work has established that this low-resolution approach has led to the mis-categorization of whole clades. For example, the Type X clade was initially reported as part of the Type II clonal lineage but was later re-categorized as a separate clonal lineage based on a single typing marker (GRA6) (Khan, Dubey, et al. 2011). Further, the low-resolution markers also failed to predict all of the genetic diversity that can be identified among natural isolates of both Type I (Khan, Behnke, et al. 2009) and Type II strains (Verma et al. 2015; Parameswaran et al. 2010) which serves as the

basis for the in depth, genome-wide analyses that were pursued in this thesis. Our work sequencing the whole genomes of Type II and X strains (Chapter 2 and 3 respectively) has demonstrated that sexual propagation is occurring within these populations more frequently than previously predicted (Lorenzi et al. 2016; Minot et al. 2012; Shwab et al. 2014; Sibley and Ajioka 2008; Su, Zhang, and Dubey 2006). Further, the increased resolution provided by whole genome sequencing is redefining the population genetic structure of *Toxoplasma*. Chapter 3 established that the Type X clade is comprised of a group of progeny derived from one or more sexual crosses between a Type II strain and an unknown γ/δ lineage. In comparison, the Type II clade was found to be comprised of closely related strains which had recently diverged from a single lineage. However, while these strains displayed genetic diversity within their genomes, SNPs were not randomly distributed as would be expected for strains that are expanding exclusively asexually. Rather, distinct haploblocks of genetic diversity within the Type II clade displayed evidence of intra-clade/unisexual recombination, as described in Chapter 2. Because unisexual expansion is virtually indistinguishable from asexual expansion at low-resolution typing, the genetic basis for the maintenance of highly successful clones within the *Toxoplasma* population structure requires analyses to be conducted using high resolution genomic typing in order to distinguish the relative contribution of each of these replicative cycles in maintaining the clonal structure in North America and Europe.

While unisexual mating appears to play a central role in maintaining *Toxoplasma's* clonal population structure, outcrossing between genetically diverse strains is key to the expansion of this generalist parasite's biological potential and it ability to infect a broad range of intermediate hosts. For example, genetic hybridization enables avirulent haploid parents to produce virulent progeny when different virulence genes are recombined into novel associations that alter the

parasites fitness or capacity to cause disease (Grigg, Bonnefoy, et al. 2001). Experimental crosses have identified a series of genetic factors that are maintained cryptically across the parasite's genetic diversity. For example, the $ROP5_{I/III}$ allele was identified to be a virulence factor only when it was combined with the $ROP18_{I/II}$ allele (Niedelman et al. 2012; Saeij et al. 2006). In Chapter 3, infection of mice with different Type X strains identified a range of virulence phenotypes among the natural population of sexually derived hybrids studied. Because Type X strains did not possess allele combinations of known virulent genes that predicted their murine virulence phenotypes, it was hypothesized that the mouse virulent Type X clones possessed novel virulence factors. To test this hypothesis, a series of sexual crosses were attempted between the mouse virulent Type X strains 3142 and 3133 and the mouse avirulent Type II B7 (ME49) clone. Only a few progeny could be recovered from the several independent sexual crosses that were attempted. Although too few recombinant progeny were obtained to perform a forward genetic QTL to identify putative virulence loci, a novel method was developed to determine the relative rates of self-mating versus outcrossing among the hybridizing strains.

During the process of isolating cross progeny, subsequent analyses performed on the recovered progeny identified several limitations to the traditional approach used for experimentally isolating recombinant progeny. The traditional method involved enzymatic/mechanical release of sporozoites from sporulated oocysts that are doubly-drug selected to identify recombinant progeny within the population. This technique cannot determine the relative rates of self-mating versus outcrossing that occurs during the sexual cycle due to the selection biases inherent in the process. To more accurately assess the frequency with which self-mating is occurring within the sexual cross, sporozoites were hatched from oocysts in a manner

that facilitates direct cloning prior to drug selection. To do this, macrophage-facilitated limiting dilution isolation was used to increase effective yields and generated significantly more cross progeny in less time using a safer isolation process than traditional excystation. Additionally, these improvements allowed for the isolation of individual progeny from the cross significantly reducing biases present in previous isolation methods. Due to these improvements, our results established that outcrossing between two parents is rare compared to self-mating within a definitive feline host, which provides a possible explanation for how these sexually fecund parasites can maintain a clonal population structure in nature.

Results

Traditional Excystation Methods Limit the Genetic Analyses of Sexual Crosses

Forward genetic approaches were employed to generate sexual cross progeny by backcrossing the mouse virulent Type X strains 3133 and 3142 with the avirulent Type II strain ME49 (clone B7 with is FUDR resistant) in order to identify novel virulence alleles within the Type X clade. In order to perform the genetic crosses, drug resistant isolates for each Type X cross parent were first generated via ENU-mutagenesis, followed by single drug (sinefungin) resistance selection. Resistant isolates were cloned, then used to produce bradyzoite cysts in the brains of mice. Sulfadiazine treatment was used to allow murine survival of acute infection of the virulent Type X strains 3133 and 3142 at a dose empirically determined to maximize murine cyst burden and survival. These cysts were fed orally to cats to allow both parental strains to undergo gametocytogenesis and fusion during the sexual phase of the life cycle. Oocysts shed from feline

feces were then collected, sporulated and excysted to generate sporozoites (Dubey and Frenkel 1976) to be screened for the presence of recombinant progeny.

Using traditional excystation protocols (commonly referred to as oocyst hatching), oocysts are extensively manually manipulated, a process which typically takes 8-10 hours to complete. Once oocysts are excysted into sporozoites, they are plated as a pooled sample into a single flask of adherent foreskin fibroblasts where the sporozoites undergo competitive infection and growth cycles as a pooled population. The pooled sporozoites are then subjected to further growth pressure by double-drug selection to select for progeny resistant to both parental drug selections, and to thereby select for recombinant progeny. Individual progeny can then be isolated, but only after the host cell lysis has occurred under double-drug selection, which typically takes several weeks after the initial excystation. Several biases were observed in our results when using this traditional excystation method. The competitive growth of all excysted sporozoites in a single pooled sample created a *de facto* competitive growth assay in which faster growing sporozoites were over-represented in the pool. The competitive growth observed biased the derived pool of sporozoites from the initial pool of recombinant progeny. This was determined by PCR-RFLP analysis after single cell cloning which established that a single, dominate clone had expanded in the Cross 1 Hatch 1 pool from the 3133XME49 cross, whereas PCR-RFLP analysis of the non-drug selected pool as a whole showed a distinct pattern of sequence types (data not shown). Furthermore, double-drug selection biased the pool toward only those progeny that had recombined to inherit a drug resistance gene from each parent and excluded analysis of those progeny that had undergone recombination but did not inherit both drug resistance loci. In order to better interrogate the individual progeny from a sexual cross and to infer the ratio of uniparental versus outbred progeny that derived from this genetic

hybridization, several modifications were made to allow for an unbiased isolation of individual progeny.

Limiting Dilution Isolation Removes Growth Biases and Macrophage-Facilitated Limiting Dilution Improves Yield of Individually Isolated Sporozoites

Significant modifications to the traditional excystation method allowed for safer, more efficient, and unbiased selection of independently isolated sporozoites (Table 1A). In order to overcome the competitive growth bias that double-drug selected sporozoites experience, a limiting dilution isolation method was established. Immediately following excystation, sporozoites were plated into successively diluted, individual wells of host cells over several 96-well plates in order to isolate individual progeny and thereby remove any implicit bias introduced by competitive growth in a pooled population. Wells were then interrogated after ten days of growth by plaque assay to identify wells that derived from a single sporozoite founder. Individual progeny were replica plated and analyzed under non-selective medium, single, or double drug selection to determine which individual progeny expanded under the various selection conditions and were thus probable recombinant progeny from the sexual cross. A number of individually isolated progeny were then selected for whole genome sequencing and further genetic characterization. The isolation of sporozoites immediately following oocyst excystation allowed for significant improvement, by at least 28%, in individual progeny isolation because immediate isolation eliminated the competitive growth between sporozoites that biased them toward those which grow best in culture (Table 1A). In the limiting dilution isolation method, all sporozoites isolated were independent progeny from the sexual cross rather than identical clones of the same progeny that exhibited a growth advantage under pool selection. In

addition to limiting dilution, a further improvement increased yields by utilizing a natural oocyst excystation method which improved safety over the traditional isolation method. This new technique was based on a recent study that examined murine subcutaneous oocyst infection and determined that systemic *Toxoplasma* infection of the murine host is facilitated by macrophage phagocytosis and excystation of infecting oocysts (Freppel et al. 2016). Although the mechanism by which macrophage phagolysosomes excysted oocysts *in vivo* was unclear, these experiments also showed that cell cultured, *in vitro* macrophages excyst oocysts and suggested that macrophage excystation could be utilized to excyst oocysts to produce viable sporozoites in culture. To overcome the low yields of sporozoites obtained by traditional chemical, enzymatic, and manual excystation methods, bone marrow derived macrophages were utilized to promote excystation of the sporozoites from sexual cross oocysts. Briefly, oocysts were incubated with cultured macrophages overnight to break the oocyst outer wall. Compared to manual methods, macrophage-facilitated excystation can be completed within 3 hours, rather than the 8-10 demanded by traditional excystation. Importantly, the sporozoite yield derived from macrophage excysted, limiting dilution isolated crosses was significantly increased by at least 290% above those sporozoites yielded from traditional methods (Table 1A).

To compare the improved isolation methods with traditional excystation methods, an equal oocyst starting concentration from a single cross of 3133 SNF^R and ME49 $FUDR^R$ was hatched in parallel using both traditional and macrophage-facilitated oocyst excystation methods, followed by limiting dilution isolation. The resultant plaques were enumerated after ten days as a measure of sporozoite yield (Table 1A). Following traditional excystation, sporozoite plaques were found in 114 wells, 93 of which were singleton individual sporozoite isolates. In contrast, following macrophage-facilitated excystation, 250 wells contained too many sporozoite plaques

to count. In the 194 wells derived from macrophage-facilitated excystation that contained a countable number of sporozoite plaques, 119 contained singleton plaques of individual sporozoite isolates. This method produced a 290% or greater increase in the sporozoite progeny pools based on the number of growing wells (which underestimates the number of excysted sporozoites per well) after excystation. The true rate of recombination within a single sexual crossing event was next interrogated. Previously, the presence of recombinant progeny was estimated using only a limited number of low-resolution marker loci characterized from double drug-selected pools of excysted sporozoites. In this study, whole genome sequencing of individually isolated progeny was used from the sexual cross to empirically determine sexual cross recombination rates, which is not possible by the traditional excystation and pool method under double drug selection.

Sexual Recombination is Biased Toward Uniparental Mating

For the four crosses pursued as part of this thesis, three used the parents 3133 x ME49, and the fourth used the parents 3142 x ME49. For two of the four crosses, progeny were isolated in 4 separate hatching events. For each excystation, DNA was extracted from the sporozoite pool prior to single cloning. Sporozoite pool DNA was analyzed to determine the dominant parental genotype within each hatched pool (Supplemental Table 1). Traditionally, sporozoite pools that have undergone successful sexual recombination contain the presence of both parental alleles within the sporozoite pool immediately following excystation, but this is by no means a quantitative measurement, and is rather a binary observation that both parents produced oocysts within the single cat infection. Observation of both parental alleles in the pool cannot distinguish between both parental strains undergoing self-mating and successful sexual recombination

resulting in the production of recombinant progeny containing an admixture of parental alleles in the same genome. In previous studies, when only a single parental allele is detected at characterized loci by this low-resolution methodology, the cross was typically abandoned, under the assumption that only one parent had undergone sexual expansion by self-mating. In this study, the PCR-RFLP characterizations of the sporozoite pools were dominated by a single parental lineage and displayed no signs of recombination between the parents at any loci interrogated (data not shown). However, the dominant parent for each of the four crosses was not biased toward one particular parent across independent replications of these crosses using the three parental strains. Of the four crosses, the PCR-RFLP analysis identified that 3133 was the only parent detected for both Cross 1 and 2 of the 3133 x ME49 cross, whereas ME49 was the only parent detected for Crosses 3. ME49 was the only parent detected for the only cross (Cross 1) done between 3142 x ME49. Overall, an equal segregation of the parental types was found to have expanded by self-mating within the Type II x Type X sexual crosses, with 2 Type X and 2 Type II strains dominating the genetic populations across the 4 cat experiments performed (Table 1C). For each cross, independently isolated progeny were grown up and whole genome sequenced in the absence of drug selection. In all cases, the dominant parent in the sporozoite pool matched with the majority of individual progeny isolated (Supplemental Figure 1 and 2). Multiple factors may induce preferential self-mating such as growth rates of each parent, the viability of the parental bradyzoite cysts, and the capability of each strain to differentiate into merozoites and commit to gametocytogenesis in the feline gut. However, neither parental strain grew significantly better than the other in the murine hosts and all three parental strains were capable of producing viable oocysts by self-mating in cats. In fact, the dominant parental inheritance of each sexual cross was not as expected based on the *in vitro* culture of the

bradyzoite cysts used to produce the crosses. Although the bradyzoites used to generate the crosses were fed to the cats in approximately equal number, cell culture viability of the parental bradyzoites was not predictive of the dominant parental genotype present in the sporozoites. For instance, although the sporozoite pool from Cross 3 of 3133 x ME49 was dominated by the ME49 parental alleles, only the bradyzoites from the 3133 (non-dominant) parent grew in culture (data not shown).

To verify these observed sexual recombination biases, each cross was repeated in multiple felines. Additionally, oocyst excystation was repeated to encompass the true diversity of progeny within the sexual cross. Thirty-one progeny from the 3133 x ME49 cross and 48 progeny from the 3142 x ME49 cross were individually isolated by limiting dilution immediately following excystation. To characterize progeny genotypes, each isolate was sent for whole genome sequencing. Although PCR-RFLP analysis of pooled sporozoites from both crosses indicated that self-mating of only a single parent had occurred (Table 1B), whole genome sequence analyses of individually isolated progeny indicated otherwise, that each of the sexual crosses yielded recombinant progeny, but with a low frequency (Supplemental Figure 1 and 2). The identification of recombinant progeny in each cross indicated that the parental strains did successfully undergo sexual outcrossing. Sequencing of isolated progeny indicated that the meiotic genetic recombination of progeny occurred at a much lower frequency (3.2% and 2.1%) (Table 1C) than previously published for sexual crosses (Pfefferkorn and Pfefferkorn 1980), which had a recovery frequency of 12.5%. In fact, traditional excystation with PCR-RFLP marker analysis would have deemed these crosses as failures, as the dominant parental strain from the initial sporozoite pool had appeared to preferentially self-mate, and none of the progeny would have been analyzed to determine if they were genetic hybrids between the two parents.

However, the work herein indicates that low rates of sexual recombination between two parents is occurring within largely self-mated sexual progeny pools.

A

3133 SNF X ME49 FUDR										
Cross	Hatch	Method of Hatch	Cross Date	Hatch Date	Well Plated	Well Lysed	Plaque Wells	Multi-plaques	Singletons	Total Plaques
1	1	Pool to Isolates	5/1/14	6/5/14	1056	0	112	16	105	121
1	2	Plate across 1 96-wells	5/1/14	6/19/14	80	0	80	647	0	647
1	3	Plate across 10 96-wells	5/1/14	5/27/15	864	0	9	0	9	9
1	4	Serial Dilution 4 96-wells	5/1/14	7/25/15	384	0	14	4	12	16
2	1	Serial Dilution 5 96-wells	12/22/15	1/11/16	480	96	384	2036	3	2039
3	1	Standard Serial Dilution 10 96-wells	11/1/16	12/15/16	960	0	141	114	93	207
3	1	Macrophage Serial Dilution 10 96-wells	11/1/16	12/15/16	960	250	194	206	119	325
3142 SNF X ME49 FUDR										
Cross	Hatch	Method of Hatch	Cross Date	Hatch Date	Wells Plated	Wells Lysed	Plaque Wells	Multi-plaques	Singletons	Total Plaques
1	1	Pool to Isolates	5/1/14	6/5/14	960	0	299	120	239	359
1	2	Plate across 1 96-wells	5/1/14	6/19/14	84	0	84	334	7	341
1	3	Plate across 10 96-wells	5/1/14	5/27/15	96	0	1	0	1	1
1	4	Serial Dilution 4 96-wells	5/1/14	7/2/15	384	0	76	12	70	82

B

Parents	Cross	infected tz	Type II bz	Type X bz	sz bias
3133 x ME49	1	1000	unknown	3000	3133
3133 x ME49	2	1000	375	0 apparent	3133
3133 x ME49	3	1000	0 apparent	0 apparent	ME49
3142 x ME49	1	1000	unknown	2800	ME49

C

	3133XB7 Cross	3142XB7 Cross
B7 Selfing	0	47
Type X Selfing	29	0
Recombinant Progeny	1	1
Total Progeny	30	48
% Recombination	3.33%	2.08%

Table 4-1: Improvements to oocyst hatching and recombination frequencies of Type X by II crosses.
A) Oocyst Hatching Improvements. Oocyst excystation of the 3133 SNF^R x ME49 FUDR^R and 3142 SNF^R x ME49 FUDR^R crosses. Oocysts from 4 independent feline crosses were hatched at least once. Each Hatch represents a separate oocyst excystation. Method of hatch corresponds to procedures listed in the methods section. Wells plated indicate the number of 96-well plates used to seed the excysted progeny. Well lysed indicates the number of wells with either complete lysis of the host cells at the time of counting or containing plaques that were too numerous to count. These wells were not included in plaque well counts. Plaques were counted in singleton plaque wells and multi-plaque wells to determine the number of independent progeny derived from these methods. Cross 3, Hatch 1 of 3133xME49 was hatched in parallel by both traditional and macrophage-facilitated excystation. Macrophage excystation yielded more growing wells of sporozoite progeny and more total countable plaques than the traditional excystation method. **B)** Parental biases enumerated across the sexual crosses. Sexual crosses from A were assayed by PCR-RFLP before and after sexual recombination and parental *Toxoplasma* contribution was identified. For each parental strain, 1000 tachyzoites (tz) were infected into murine hosts. The resultant bradyzoites (bz) were counted before sexual recombination occurred. Sporozoite (sz) bias was determined based on PCR-RFLP of both pooled sporozoites and unselected individually isolated progeny. **C)** Recombination frequencies based on whole genome sequencing of progeny from two distinct Type II by Type X crosses. Verified recombinant progeny are determined based on SNP density analysis (See supplemental Figures 1&2).

Discussion

Improved Methods of Excystation Also Improve Interrogation of Recombinant Progeny

Sexual recombination rates between parental strains were interrogated using independently isolated progeny from a sexual cross between the Type II and X clades. There are several steps of the excystation process that introduce biases affecting the isolation of recombinant progeny, and subsequently the true ratios of sexual recombination events that occur in the feline host. The discrimination between true ratios of outcrossing versus self-mating was resolved between the parental strains by improving the isolation method that eliminated several experimentally induced biases such as competitive growth and survival of manual isolation. Independently isolated sexual cross progeny derived using these methods allowed for the unbiased interrogation of sexual cross recombination rates. In addition to the computational power of these new excystation methods, the safety of excystation of oocysts was greatly improved. Previous methods of oocyst hatching required both specialized reagents and lengthy manipulation of highly infectious oocysts (Freyre and Falcon 2004; Khan et al. 2005). While handing the highly infectious oocysts, researchers must remain in full-body personal protective equipment and keep equipment and generated waste isolated to reduce the risk of laboratory and personnel contamination. Human virulence phenotypes of newly isolated progeny are unknown and cannot be readily predicted, as even avirulent strains have been shown to produce virulent progeny (Grigg, Bonnefoy, et al. 2001). Thus, it is best to limit potential exposure to oocysts as the methods described here do. The length of exposure to the oocysts was more than halved by utilizing bone marrow derived macrophages (BMDMs) to digest oocyst walls, and the recovery

of viable progeny was increased exponentially, facilitating single cell cloning to increase yields of cloned progeny for molecular characterization (Freppel et al. 2016).

Although this macrophage-facilitated method may bias derived progeny toward those sporozoites that survive BMDM phagocytosis, macrophages are thought to be the preferred cellular host of *Toxoplasma* infection (Jensen et al. 2011; Melo et al. 2013; Suzuki et al. 1988). Additionally, macrophage-facilitated excystation combined with limiting dilution isolation produced, at the very minimum, a 28% increase in the number of individually isolated sporozoites available for interrogation. However, it is of note that individual progeny isolates are more heavily influenced by correct limiting dilutions than on any improvement in technique as wells with multiple plaques indicated that far more sporozoite progeny were obtained by the macrophage excystation technique than could be enumerated here. Thus, the true improvement using macrophage-facilitated excystation followed by limiting dilution of progeny from oocysts demonstrated in this study has significantly improved yields of independent sporozoite progeny isolates by at least 300%. In previous genetic hybridization studies recovery of between 20-35 recombinant progeny was necessary to determine novel virulence alleles (Behnke et al. 2011; Behnke, Khan, and Sibley 2015; Saeij et al. 2006; Taylor et al. 2006). With the increased numbers of isolated progeny herein, even if some bias is shown in the derived progeny, a single hatch yields ten times more individually isolated progeny than was necessary for prior studies to characterize novel virulence alleles. Additional isolates have also allowed the distinction between low-frequency recombination in a sexual cross and sexual self-mating during outcrossing co-infection.

Biased Sexual Recombination is Common in Protozoan Parasites

Macrophage-facilitated limiting dilution excystation permitted the examination of all individually isolated progeny derived from a single cross. Characterization of these progeny by WGS (Supplemental Figures 1&2) demonstrated that most progeny (96.8% in the 3133 by ME49 and 97.9% in the 3142 by ME49 cross, Table 1C) were products of uniparental mating rather than outcrossing. These uniparental mating rates are extremely high in comparison to previously reported successful crosses, in which a high-frequency of genetic hybridization was reported and in which the crosses experienced rates of uniparental mating around 50% (Pfefferkorn and Pfefferkorn 1980). The disparity between the experimental crosses reported here and previously reported successful crosses is likely due to the elimination of progeny selection biases in our interrogation. A common recombination verification procedure uses PCR-RFLP to interrogate the pool of oocysts prior to excystation and the pool of excysted sporozoite progeny immediately following excystation to verify that sexual recombination has occurred. If biallelism or marker recombination is not observed within either pool, no sexual recombination is considered to have occurred, and no further analysis is done on the oocysts from that cross. However, as seen in this study, the traditional protocol lacks the inherent resolution to detect low-frequency recombination masked by highly prevalent self-mating. In fact, all crosses interrogated here would have previously been classified as failures. As a result, the number of successful crosses between *Toxoplasma* strains has likely been underestimated. It was demonstrated here that even crosses initially thought to be failures contain both low-frequency genetic progeny from outcrossing within a larger population of progeny derived from uniparental self-mating by one of the parents. Thus, genetic hybridization is likely more common than previously reported as many unreported crosses are characterized as failures, when only one parent appeared to have

expanded by uniparental mating. Despite previous reports from the analysis of successful sexual crosses within the *Toxoplasma* field, the potential for a parasite to preferentially self-mate is not a rare occurrence (Boothroyd 2009; Grigg and Sundar 2009; Minot et al. 2012; Su et al. 2003; Wendte, Miller, Lambourn, et al. 2010). Such a bias toward self-mating has also been seen in *Cryptococcus neoformans* as well as other protozoan parasites, including *Leishmania* where hybrid progeny only accounted for 14% or less of all progeny isolated (Billmyre et al. 2014; Sun et al. 2014; Inbar et al. 2013). For *Toxoplasma,* this uniparental bias within sexual mating has been enumerated and defined in the studies herein for the first time.

Host and parasite factors are likely co-operating to promote uniparental mating. The bradyzoite cysts used for feline infections have variability in the number of parasite bradyzoites contained within a single host tissue cyst and have variable viability across different host-strain combinations. This variability has the potential to affect parasite numbers that can participate in sexual recombination within the definitive host. Additionally, differential seeding capacity of parasites inside the feline intestinal tract has been observed in other parasites, such as the gregarine apicomplexans, *S. terebellae* and *S. melongena n. sp*. (Tenter 1995; Wakeman, Heintzelman, and Leander 2014). *Toxoplasma* is known to display differential host partitioning during co-infections of more than one *Toxoplasma* strain, which segregate differentially throughout the bodies of their hosts (Verma et al. 2017). In the feline host, bradyzoites may have a strain-specific ability to differentiate into gametes in different locations and with differential maturation rates inside the feline intestine. If so, gametes produced closer together both biologically and temporally would be more likely to merge into a zygote to produce oocysts. In order to characterize these differences, more controlled studies of the sexual cycle within the feline gut are necessary.

Biased Sexual Recombination Shapes the Population Structure of *Toxoplasma*

Sexual recombination in nature has been classified as a rare event for *Toxoplasma* (Sibley and Ajioka 2008; Wendte, Miller, Lambourn, et al. 2010). However, careful examination of Type II strains has indicated that sexual recombination may be hidden within lineages as a result of same-clade/unisexual mating (Chapter 2). Sexual recombination is an important mechanism by which *Toxoplasma* retains the capability to diversify its genome while still maintaining virulence genes cryptically in its genome (Wendte, Miller, Lambourn, et al. 2010). Uniparental mating may drive clonality even in cases of co-infection, where outcrossing is predicted to be prevalent. To determine if uniparental and/or highly-similar unisexual mating (mating within a clade) is preferred over sexual outcrossing, more sexual crosses would need to be examined using the new, unbiased techniques of macrophage-facilitated excystation followed by limiting dilution isolation of sexual progeny detailed here. Furthermore, there are several factors to consider in experimental crosses that may bias the outcome. For example, bradyzoite viability and associated gamete differentiation and propagation should be assessed within the feline intestine. While it is currently standard to infect a feline with approximately equal bradyzoites per strain, not all bradyzoite cysts contain the same number of bradyzoites thus accounting for this disparity may be key in the standardization of feline infections. Therefore, these bradyzoite infections in cats would also need to be induced such that gamete production from the bradyzoites of each strain is optimized to produce sexual recombination between the maturing gametes of infecting parasites. Ideally, each strain would produce gametes at the same time in overlapping locations across the feline host gut to ensure the greatest likelihood of cross-strain gametes encountering and fusing to produce the hybrid zygotes that would yield recombinant sporozoites. In addition to the possible biological limitations in these studies, feline experiments

are expensive and more difficult to control. Unlike inbred mice, the host population of felines is not as tractable to repeat infections. As a result, the significance of these factors in facilitating uniparental mating over outcrossing currently remains a black box in the *Toxoplasma* field. Thus, isolation of oocysts from natural populations as well as repeated experimental crosses with more diverse strains, such as South American strains, may be the best approach to determine the true prevalence of outcrossing versus self-mating within a natural sexual cross.

Materials and Methods

Culture and DNA Isolation of Toxoplasma Strains

Toxoplasma strains were cultured on human foreskin fibroblast (HFFs) monolayers at 5% CO_2 at 37°C. These HFFs were grown in DMEM Complete Medium derived from Dulbecco's Modified Eagle Medium (DMEM) enriched with 10% heat-inactivated fetal bovine serum (FBS), glutamine, and treated with gentamycin and penicillin-streptomycin.

Tachyzoites were grown until HFF monolayer lysis. Tachyzoites and lysed host cells were filtered using a 3.0-micron filter to remove host cellular debris. Clean tachyzoites were pelleted and rinsed with PBS. DNA was extracted from parasite pellets using the Qiagen DNeasy Blood and Tissue Kit. 1 ng of DNA was sent to Rocky Mountain Laboratories for whole genome sequencing as described below.

Macrophage Growth for Oocyst Hatching

Bone marrow derived macrophages (BMDMs) were made from 6-8 week old female C57BL/6 mice (Jackson Laboratories), as described (Trouplin et al. 2013). Briefly, bone marrow

was harvested from both mouse femurs and re-suspended in complete macrophage media (Dulbecco's Modified Eagle's Medium, 10% fetal bovine serum, glutamine, HEPES, and gentamicin) supplemented with 30% L929 cell supernatants. Cells were plated across non-tissue culture treated Petri dishes and incubated at 37°C in 5% CO_2. Media was exchanged on days 3 and 6, and adherent cells were harvested on day 8 by scrapping with a rubber policeman. Cells were washed with complete media and counted. Counted BMDMs were plated into five T25 flasks at 7 x 10^6 BMDMs/flask.

Initial Characterization of Progeny from Sexual Crosses

Because the previous typing markers for Type X strains do not clearly distinguish Type X and II strains via restriction digest, a PCR-RFLP scheme was created to screen the progeny from the II by X sexual cross. To characterize recombinant progeny from the sexual crosses between the Type II and Type X parents, PCR-RFLP typing markers at BSR4, L358, GRA7, BAG1, SAG1, GRA6, ROP1, and SRS9 were developed to distinguish between the Type II and Type X alleles using the comparison of nucleotide sequences of typing markers known in our laboratory (Supplementary Table 1). Restriction enzymes were chosen that distinguish the Type II and X parental strain sequence at the chosen typing markers to create a PCR-RFLP map of progeny from the sexual crosses. Type X strains and sexual cross progeny were also characterized using these PCR-RFLP typing markers to allow the differentiation of the 3133 and 3142 Type X strains both from each other and from the ME49 Type II strain. Markers were amplified by PCR using Taqman Ampli-Taq polymerase and an Eppendorf thermocycler as previously described (Parameswaran et al. 2010; Su, Zhang, and Dubey 2006). PCR products were restriction enzyme

digested using the enzymes listed in Supplemental Table 1 for 1 hour at 37°C and products were run on a 1% agarose gel and compared to parental strain digest to verify marker identity.

Characterization of Whole Genome Sequencing

1 ng of DNA from DNA extractions of each isolate was sent to Rocky Mountain Laboratories for whole genome sequencing using Illumina HiSeq technology. Fastq reads were reference mapped against the ToxoDB version 8.2 ME49 genome using BWA 0.7.5a to align the reads to the reference genome and GATK 3.7 in coordination with Picard 1.131 following best practices to quality control the mapped reads (Van der Auwera et al. 2013). Following mapping, the gVCF method of GATK was used to combine SNP calls using stand_call_conf of 30.0, nct of 10, and ploidy of 1 (for haploid genomes) to call 568592 single nucleotide polymorphism positions across the whole genomes of these strains (Van der Auwera et al. 2013).The derived VCF formatted SNP file was curated using GATK and VCFTools to produce a tabular file containing only biallelic SNPs which excluded large insertions or deletions (Danecek et al. 2011).

The tabular, biallelic SNP file was analyzed using a custom R script to display strain-specific SNP locations (As seen in Chapters 2 and 3). SNPs were binned into 100 kbp windows across the genome(Minot et al. 2012; Yin, Cook, and Lawrence 2012). The larger the number of SNPs, the more divergent the strain was from the Type II ME49 reference strain.

Virulence Gene Characterization of Type X Strains

Whole genome sequence fastqs derived above were reference mapped to known virulence genes within the *Toxoplasma* genome. The known sequences of virulence genes of ROP5, ROP18, ROP16, ROP17, and GRA15 were used to reference map the Type X strains (Saeij et al. 2006; Saeij et al. 2007; Shwab et al. 2016; Taylor et al. 2006; Rosowski et al. 2011). SNPs within these strains were characterized as above, and using GATK and BEDTools, consensus sequences were called to create fastas of these known virulence alleles (Quinlan and Hall 2010; Van der Auwera et al. 2013). These fasta files were compared to one another using SeqMan Pro 14.0 and polymorphisms were used to characterize the allelic inheritance of these strains.

Sexual Crossing in Cats

N-ethyl-N-nitrosourea (Sigma-Aldrich) was used to derive mutant pools of drug resistant *Toxoplasma* from the Type X 3133 and 3142 strains. Mutagenized pools of Type X strains were selected for sinefungin resistance, following published protocols, to create 3133 SNF^R and 3142 SNF^R (Khan et al. 2014; Behnke, Khan, and Sibley 2015). The Type II ME49 $FUDR^R$ strain has previously been published and was obtained from the Sibley lab (Khan et al. 2014). Cloned drug-resistant strains were intraperitoneally infected into 8-week old female outbred CD-1 mice from Charles Rivers Laboratories at a dose of 1000 tachyzoites per mouse. To ensure murine survival, Type X infected mice were treated with 0.5 mg/ml of sulfadiazine (Sigma-Aldrich) in their drinking water on days 5-15 post-infection (Sibley 2009). Mice were monitored for survival and cachexia daily for 42 days. Serum was collected on day 20 post-infection and seroconversion was tested via IFAT (Miller et al. 2002). 42 days after infection, infected mice were euthanized,

and their brains were removed. Brains bearing bradyzoite cysts were resuspended in 1ml of sterile PBS and passed through 19G and 21G needles to homogenize. 20 microliters of homogenate containing bradyzoite cysts were counted on a slide schemer. Type II ME49 $FUDR^R$ and Type X SNF^R brain homogenates were combined in approximately equal ratios. The mixed brain homogenates were sent to the USDA Dubey lab where they were fed to feline hosts for sexual recombination. Feces were collected from felines and oocysts separated via floatation in sucrose and sporulated and stored in sulfuric acid as previously published (Dubey 1995, 2001).

Excystation of Sporozoites from Oocysts

Oocyst can be stored at 4°C for several months in sulfuric acid, however the viability of the excysted progeny is severely reduced the longer the oocysts are stored. Thus, immediate oocyst excystation is recommended as soon as possible following oocyst isolation from feline feces. As oocysts are highly infectious and sexual recombination is known to increase virulence in murine hosts, personal protective equipment (double-gloves, face mask, safety glasses, disposable lab smock, and shoe covers) should be worn, lockable lids for centrifuge buckets used, work done in a biosafety hood, and hoods should be protected with disposable liners to reduce the risk of laboratory contamination. After all work is completed, sterilize all equipment with 70% ethanol, followed by fungicide, followed by 70% ethanol, and autoclave equipment which can be autoclaved.

Traditional Oocyst Excystation

Traditional oocyst excystation was developed by both the Sibley and Dubey laboratories and is detailed below (Khan et al. 2014; Dubey 1995). 20 ml of sporulated oocysts from fecal extractions were pelleted at 1300 rpm for ten minutes in a 50 ml conical tube. Supernatant was discarded, and the pellet of oocysts was washed in PBS three additional times to neutralize the oocyst sulfuric acid storage buffer. The pellet was placed on wet ice and resuspended in 5 ml of PBS. 5 ml of 10% bleach in PBS was added to the solution to sterilize the bacteria in the pelleted oocysts and the resuspended oocysts were incubated on ice for 45 minutes. Bleach-treated oocysts were diluted to 50 ml with PBS and spun at 200 rpm for 5 minutes to remove large fecal debris in the resulting pellet from the oocysts. The supernatant was moved to a new 50 ml conical tube and spun at 2000 rpm for ten minutes. The resulting pellet containing the oocysts was resuspended in 50 ml of PBS and washed three times by spinning at 2000 rpm for ten minutes. The oocyst pellet was resuspended in 1 ml PBS in a 15 ml polystyrene conical tube. 2g of 450-600 micron glass beads (Sigma G-8772) were added to the resuspended oocysts. The bead-oocyst solution was vortexed on high for two minutes to break the outer oocyst wall and beads allowed to settle via gravity. The supernatant from the bead-oocyst solution was decanted and spun at 1500 rpm for ten minutes. The oocyst pellet was resuspended in 1 ml of 5% sodium taurodeoxycholate (bile salts to induce oocyst inner wall cracking) (Sigma T0875-5G) in RPMI and incubated at 37°C for exactly ten minutes. After ten minutes, the oocyst-bile salt solution was diluted with 10 ml of ice cold RPMI. The oocyst solution was spun at 2000 rpm for ten minutes. The resulting oocyst pellet was washed three times with 15 ml of RPMI and spun down at 2000 rpm for ten minutes. The resulting pellet contained the excysted sporozoites which were resuspended in 24 ml of DMEM Complete Medium and distributed evenly across 3-T25 flasks of

HFFs. HFFs were allowed to grow for 1 week before drug-selecting. Following excystation and one week of growth on HFFs in culture, double drug selection of sporozoite parasites was done on the growth pool of sporozoites using SNF (24×10^7 M) and fluorodeoxyuridine (FUDR, 24×10^5 M) as determined based on parental drug-selection.

Limiting Dilution Isolation Following Oocyst Excystation

Individual isolation of sporozoites improves the yield of traditionally obtained progeny of oocyst excystment (detailed above) by eliminating competitive growth. Instead, individual sporozoite progeny are isolated immediately following excystation rather than after drug selection. Rather than plating sporozoites in a pool within T25 flasks (as above), the excysted sporozoite pellet was re-suspended in 48 ml of Complete DMEM Medium. Half of this re-suspension of sporozoites was plated across one 96-well plate of HFFs. The original pool was diluted back to 48 ml with Complete DMEM and the process repeated until ten 96-well plates of HFFs were plated with sporozoites, with the last plate being 256 times more dilute than the initial plate. As with a standard plaque assay, plates were allowed to grow for one week, after which individual wells were examined for plaques to identify individually isolated sporozoites. These individual sporozoites were grown until the all HFFs in the well lyse. Lysed 96-well plates were replica plated onto four 96-well HFF plates each. These growth replicates are used for single and double drug-selection (1 plate without drugs, 1 plate SNF selection, 1 plate FUDR selection, 1 plate SNF + FUDR selection) using SNF (24×10^7 M) and fluorodeoxyuridine (FUDR, 24×10^5 M) as determined based on parental drug-selection. Progeny were also characterized via PCR-RFLP and whole genome sequencing as described above

Macrophage-Facilitated Oocyst Excystation

Macrophage-facilitated excystment of oocysts has been reported to occur *in vivo* in murine subcutaneous *Toxoplasma* infections (Freppel et al. 2016). A method has been developed for *in vitro* use here to provide a safer approach that requires less manual manipulation than traditional oocyst excystation methods. Here, *in vitro* BMDMs were used to excyst oocysts. The beginning of macrophage-facilitated excystment is the same as traditional excystation methods. 20 ml of sporulated oocysts from fecal extractions were pelleted at 1300 rpm for ten minutes in a 50 ml conical tube. Supernatant was discarded, and the pellet of oocysts was washed in PBS three additional times to neutralize the oocyst sulfuric acid storage buffer. The pellet was placed on wet ice and resuspended in 5 ml of PBS. 5 ml of 10% bleach in PBS was added to the solution to sterilize the bacteria in the pelleted oocysts and the resuspended oocysts were incubated on ice for 45 minutes. Bleach-treated oocysts were diluted to 50 ml with PBS and spun at 200 rpm for 5 minutes to remove large fecal debris in the resulting pellet from the oocysts. The supernatant was moved to a new 50 ml conical tube and spun at 2000 rpm for ten minutes. The resulting pellet containing the oocysts was resuspended in 50 ml of PBS and washed three times by spinning at 2000 rpm for ten minutes. The resulting bleach-treated oocyst pellet was diluted in 48 ml of Complete DMEM and plated across 5-T25s of BMDMs (prepared as above). The BMDMs and bleach-treated oocysts were incubated together for 24 hours at 37°C in 5% CO2. After 24 hours, culture medium was removed from BMDMs and phagocytosed sporozoites and 1 ml of trypsin was added to each T25. The BMDMs and sporozoites were scraped from the plates and pooled into 48 ml of Complete DMEM Medium. The pooled BMDM sporozoites were plated as with the limited isolation procedure above until ten 96-well plates of HFFs were inoculated with serially diluted sporozoites.

Ethics Statement

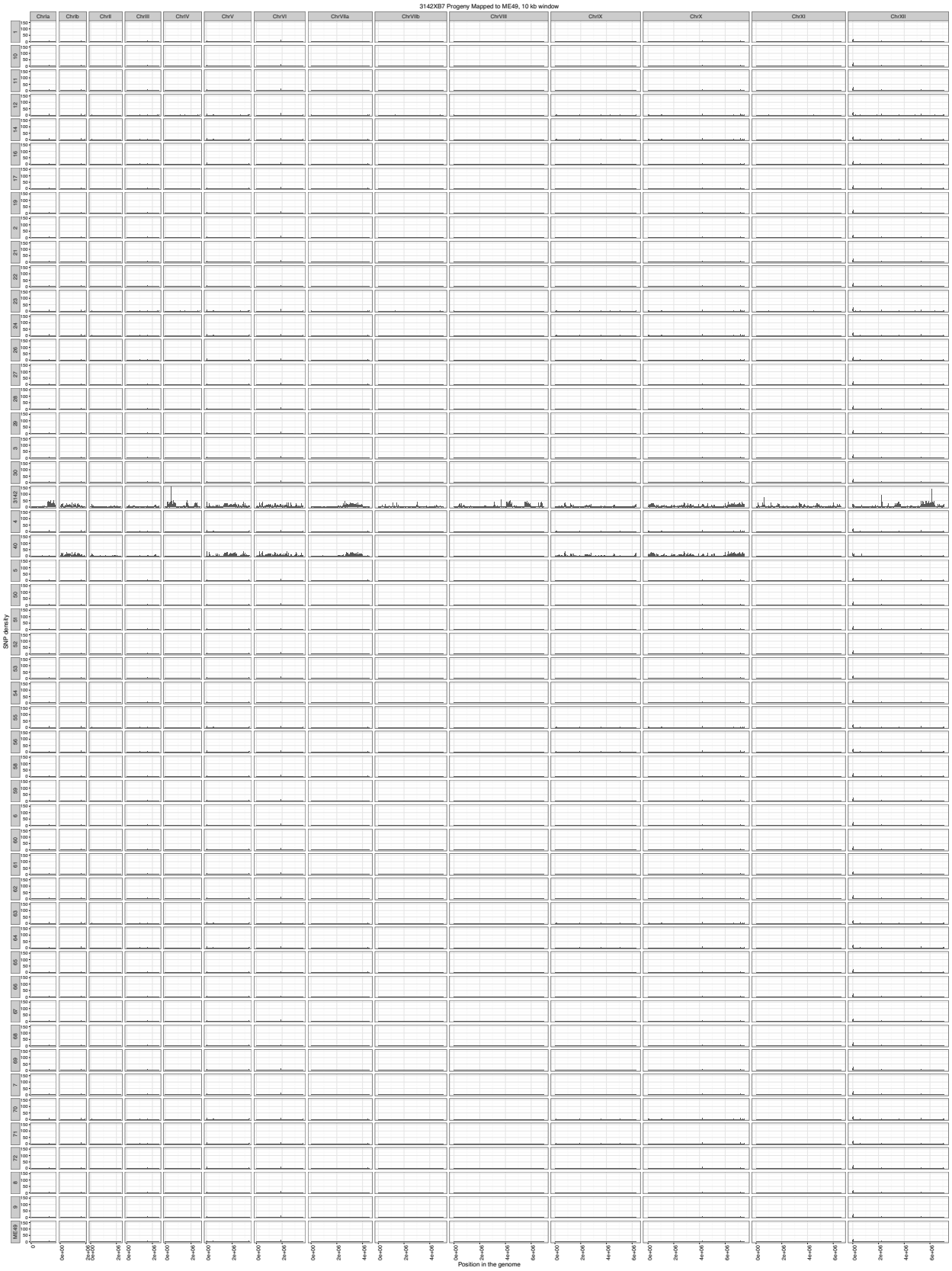
The use of animals in this study was done under the approval of the Animal Care and Use Committee of the National Institute of Allergy and Infectious Diseases, National Institutes of Health (LPD-22E). All guidelines required by the NIH and the Animal Welfare Act were strictly followed.

Supplemental Figures

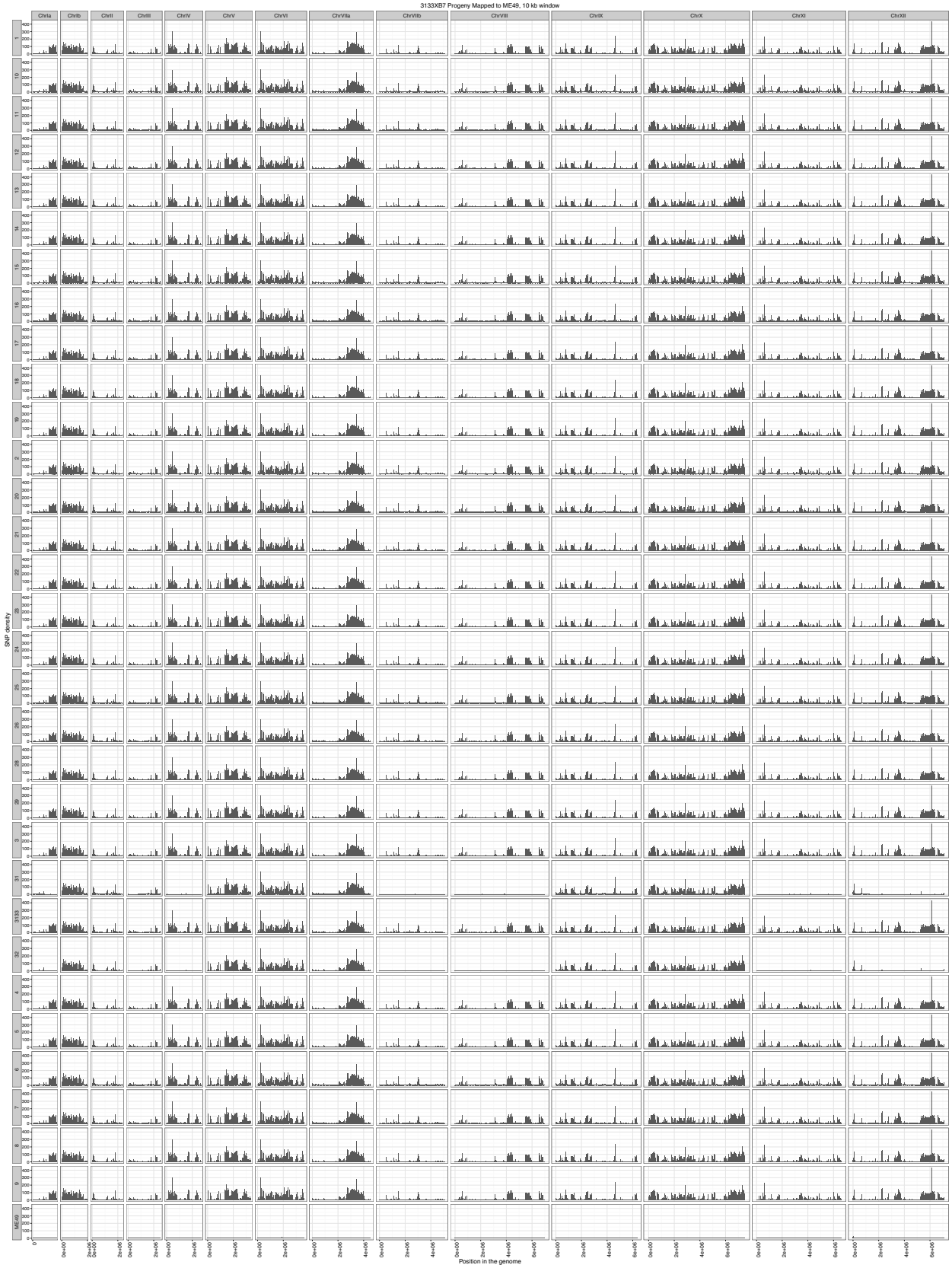
Marker	BSR4	L358	GRA7	BAG1	SAG1	GRA6	ROP1	SRS9
Chromosome	IV	V	VIIa	VIIb	VIII	X	XI	
Length	1138	600	460	1917	700	758	960	600
RFLP	NciI	HaeIII, NlaIII	EcoRI, XmnI	BbvI, AvaI, MboI	HaeIII	AvaI	DdeI	MboI
Type II	II	II	II	II	II	II	II	II
3133	X	I	II-like	X	II-like	X	X	X
3142	X	I	X	X	II-like	X	II-like	X

Supplemental Table 4-1: PCR-RFLP typing markers for the detection of cross progeny.
Markers are listed by chromosome position, sequence length and RFLP enzyme. Marker identity for each typing marker is shown based on RFLP digestion of these strains by enzymes listed in the table. Markers were selected based on the ability to differentiate parental strains based on differential fragment size of RFLP-digested PCR products.

Supplemental Figure 4-1: (Following page) - SNP density plots of all WGS progeny from the 3142 by ME49 cross.
Isolated progeny from the 3142 (Type X) x ME49 (Type II) cross were mapped to the ME49 genome and plotted in 10 kbp windows across the genome (row) in chromosomal segments (columns). SNPs per 10 kbp window are plotted on 0-150 SNP scale as vertical bars. Clone #40 is the only verified recombinant progeny from this cross.


3142XB7 Progeny Mapped to ME49, 10 kb window
ChrIa
ChrIb
ChrII
ChrIII
ChrIV
ChrV
ChrVI
ChrVIIa
ChrVIIb
ChrVIII
ChrIX
ChrX
ChrXI
ChrXII
SNP density
Position in the genome

Supplemental Figure 4-2: (Following page) - SNP density plots of all WGS progeny from the 3133 by ME49 cross.
Isolated progeny from the 3133 (Type X) x ME49 (Type II) cross were whole genome sequenced and mapped to the ME49 genome and plotted in 10 kbp windows across the genome (row) in chromosomal segments (columns). SNPs per 10 kbp window are plotted on 0-400 SNP scale as vertical bars. Progeny clones #31 and #32 are replicate clones from the same selected pool and are the only verified recombinant progeny from this cross.


3133XB7 Progeny Mapped to ME49, 10 kb window
ChrIa
ChrIb
ChrII
ChrIII
ChrIV
ChrV
ChrVI
ChrVIIa
ChrVIIb
ChrVIII
ChrIX
ChrX
ChrXI
ChrXII
SNP density
Position in the genome

Chapter 5 - Conclusions: Sexual Recombination is Common and Frequent Within and Between Clades of *Toxoplasma*

Recombination and Clonal Expansion of *Toxoplasma gondii*

Toxoplasma gondii is a generalist parasite that infects 30-80% of human populations, and 30-50% of warm-blooded animals globally and is one of the leading causes of preventable blindness worldwide (Soheilian et al. 2005; Torrey and Yolken 2013; Dubey and Jones 2008; Dubey et al. 2011; Guo et al. 2016; Lopes et al. 2014). Despite the diversity of its hosts and its fecund sexual cycle, the *Toxoplasma* population genetic structure is relatively simple. This parasite's population is dominated by a limited number of widely distributed clonal lineages, which typically cause asymptomatic infections in people and livestock (Howe and Sibley 1995; Khan, Dubey, et al. 2011; Khan et al. 2005). Genetically diverse strains are less widely distributed and are prominent in regions of the globe (such as South America) where they have been associated with outbreaks and different disease sequelae (Shwab et al. 2014; Su et al. 2012; Su, Zhang, and Dubey 2006).

Sexual recombination is prevalent across the range of parasitic life including viruses, bacteria, and other protozoans such as *Leishmania* and *Plasmodium* in spite of the risk that recombination is just as likely to separate as to conjoin beneficial gene combinations (Heitman 2010; Inbar et al. 2013; Ariey et al. 2014; Arisue and Hashimoto 2014; Miotto et al. 2013; Twiddy and Holmes 2003; Vijaykrishna et al. 2011). The sexual cycle of *Toxoplasma* has previously been shown to confer on F1 progeny an altered biological potential, such as virulence different from parental capabilities, and has been postulated to facilitate the success and expansion of the parasite's host range, due in part to the high transmissibility of its extremely

infectious oocyst stage, the end product of the sexual cycle (Grigg, Bonnefoy, et al. 2001). However, current population datasets have identified predominantly clonal population structures which seemingly contradict the expected genetically diverse population structure for a sexually reproducing organism (Lehmann et al. 2006; Shwab et al. 2014; Sibley and Ajioka 2008; Su et al. 2012). Debates within the field focus on the extent to which *Toxoplasma* propagates asexually versus sexually (Tibayrenc and Ayala 2002, 2014), whether dominant clones are the products of sexual self-mating (Wendte, Miller, Lambourn, et al. 2010) and the extent and frequency to which out-crossing produces new recombinant strains that possess altered biological potential (Grigg, Bonnefoy, et al. 2001). A primary goal of the work presented herein was to determine the frequency of sexual recombination in nature, and the role of sexual recombination in shaping the genetic population structure of *Toxoplasma*. Additionally, these WGS datasets were used to test the sufficiency of the clonal theory that currently explains the clonal population structure found in nature.

The main conclusion of this thesis is that sexual recombination plays a previously underappreciated role in shaping the population structure of *Toxoplasma*. Evidence for sexual recombination was found within previously designated clonal lineages, II and X, that predominate in North America. While the Type II clade is an ancient lineage that has undergone asexual diversification followed by unisexual (within clade) meiotic expansion, it resembles a clonal lineage when only limited markers are used for genotyping. In contrast, Type X is not a clonal lineage, but rather a clade of sexually related strains, that are also capable of expanding unisexually. Data in this thesis suggest that the Type X exists as a clade of F1 progeny from a cross between a Type II strain and a mosaic parent of two distinct ancestries (referred to as γ/δ). Combined, the Type II and X clades represent the majority of globally sampled strains. Thus,

evidence of recombination here demonstrates that sexual recombination is more prevalent than previously hypothesized, as it has been characterized to have occurred only rarely in nature (Shwab et al. 2014; Sibley and Ajioka 2008).

In addition to its role in creating variation within the genetic population structure of *Toxoplasma,* sexual replication also plays a role in the amplification of particular clones. Unisexual (intra-clade) recombination was observed within the Type II clade, characterized by the inheritance of haploblocks of shared ancestry, unique to the Type II clade, that were independently admixed among the strains, indicating that these Type II strains were capable of undergoing sexual recombination between non-identical strains within the clade, but in a biased manner. Evidence for this bias was observed during four independent sexual crosses between Type II and X parents whereby progeny derived from sexual replication were observed to strongly bias toward a single parent yet also produce, at low frequency (2-5%) recombinant progeny. This sexual bias is likely one of the major factors that masks the identification of frequent sexual replication such that *Toxoplasma* appears to bypass its sexual cycle and maintain an apparently clonal population structure. Biased sexual recombination also supports the conclusion that sexual propagation is more common in nature than previously hypothesized.

Type II Strains are Expanding Predominantly by Unisexual Mating

Type II is a globally-distributed clonal lineage that is highly successful in nature, establishing it as a model candidate clade for testing whether the clonal expansion theory holds up at whole genome resolution for parasitic protozoans (Shwab et al. 2014; Su et al. 2012; Su, Zhang, and Dubey 2006). It has previously been theorized that only asexual expansion would allow for the retention of such a widely distributed clonal population (Sibley and Ajioka 2008;

Tibayrenc and Ayala 2014). It has also been argued that sexual expansion by self-mating could likewise produce a highly clonal population structure (Wendte, Miller, Lambourn, et al. 2010). In fact, the only way to differentiate asexual expansion from sexual self-mating in this haploid parasite is to interrogate the genomes of these strains for hallmarks of sexual expansion, including changes in allelic diversity, differences in somy, genetic structural variation, and haploblock recombination. No previous study has examined a single clade of strains at whole genome resolution. All other studies have either focused on the global diversity of *Toxoplasma* strains (Lorenzi et al. 2016; Shwab et al. 2014; Su et al. 2012; Su, Zhang, and Dubey 2006) or have analyzed strains within a clade using only a limited suite of low-resolution markers that are unsuitable for genome-wide characterization because they fail to capture the true extent of diversity within a clade (Minot et al. 2012; Sibley et al. 2009; Su, Zhang, and Dubey 2006). This thesis presented an unprecedented, in-depth WGS examination of a single clonal clade of strains isolated from across multiple continents and host species.

Type II strains were compared using whole genome sequencing (WGS). Geographically distinct isolates of Type II were compared by marker, chromosomal, and whole genome sequencing and analyzed by phylogeny to compare Type II strains to other known clades of *Toxoplasma.* Although Type II strain-specific SNPs were relatively rare, allelic diversity was higher than would be expected for a clonal lineage. Based on phylogenetic comparisons and by SNP diversity mapping, a number of distinct haplotypes were found within the Type II clade. Importantly, all of these haplotypes were distinguished based on previously unknown, large, and phylogenetically distinct blocks of alleles of distinct ancestry that could be identified within Type II strains from different geographies. Importantly, several North American Type II strains were unique genome-wide admixtures that possessed haploblocks of sequence that originated

within either the North American Type II lines, or African Type II lines. Because asexual replication accumulates SNPs across the genome randomly, due to polymerase error during replication, these SNPs are generally evenly distributed across the genome rather than found in localized distinct haploblocks that track with a specific geographic region, which was observed among several of the Type II strains studied herein. Shared haploblocks between strains with clear crossover points between them is a hallmark of sexual recombination and this was found to occur among several of these hybrid strains within the Type II clade. Supporting evidence for sexual recombination was observed by the incongruence of phylogeny between the nuclear and apicoplast genomes within the Type II strains investigated. This only occurs when nuclear genomes undergo recombination during meiosis and the phylogeny of the maternally inherited apicoplast genomes (i.e., grouping with African strains) does not match with the phylogeny of the entire nuclear genome (i.e., grouping with North American strains). This evidence led to the conclusion that Type II is by no means a clonal population, but rather a closely related clade of strains that have unisexually recombined with one another.

Although outcrossing has previously been reported to occur only rarely among natural isolates (Minot et al. 2012), there is abundant evidence that Type II has recombined with several other distinct ancestries to produce both the Type I and Type III strains (Boyle et al. 2006; Minot et al. 2012). In this thesis, the Type III strain VEG was sequenced and it was shown to be an admixture of at least two distinct Type II strains that introgressed into the Type III genome, one that was highly similar to the North American ME49, the other more similar to the European strain PRU. Thus, work herein has established that recombination is occurring among distinct haplotypes within the Type II clade. Further, at least three distinct Type II haplotypes have been identified to be highly prevalent within North American Type II strains. Indeed, sequence

analysis of the various Type II haplotypes found in North American strains indicates that several strains have intermixed with African Type II strains within the same clade. This recombination was previously missed because these strains were not interrogated against one another at whole genome scale. What has not been established is whether Type II strains are biased in their mate preference during sexual expansion. Finding Type II admixed into distinct genetic ancestries in Type I, III and X strains indicate that Type II lines are capable of outcrossing. But equally, the identification of intra-clade admixtures within the Type II clade indicates that Type II can also expand unisexually, rather than strictly via self-mating. Hence, experiments needed to be set up to investigate whether *Toxoplasma* "prefers" to self-mate in the presence of an extra-clade mating partner, which would indicate whether the recombination frequency of Type II crosses biases toward either self-mating or unisexual mating between a highly similar, same clade parental strain, which would allow for the expansion of a closely-related "clonal" population. In Chapter 4 experiments were set up to determine the ratio of outcrossing to self-mating in order to test whether biased mating is likely occurring in nature.

The demonstration of high levels of sexual replication inherently disputes the clonal theory proposed to explain the clonal population dynamics of *Toxoplasma*. Furthermore, the genetic diversity within the Type II clade, the most prevalent of the presumed clonal clades, suggests that previous population studies have underestimated global genetic diversity, and failed to address unisexual mating as an alternative mechanism to generate seemingly clonal genetic population structures within the parasitic protozoans. The asexual expansion theory is predicated on utilizing only low-resolution typing markers to characterize *Toxoplasma* strains. Further, these PCR-RFLP typing markers are biased, as they were designed to distinguish only between SNPs that vary between the three major clonal lineages (I, II, and III). Thus, they underestimate

to the total diversity inherent within the genetic population structure of *Toxoplasma* (Su, Zhang, and Dubey 2006). Recombination of the unique Type II haploblocks between the strains of the Type II clade indicated that it was necessary to test whether intra-clade sexual replication was extant across the entire population structure of *Toxoplasma*. To further interrogate the extent of intra-clade/unisexual recombination across the population the analysis was extended to another clonal lineage, referred to as Type X, or HG12, and reported in Chapter 3.

Type X Strains Resemble F1 Progeny from a Natural Sexual Cross

The release of oocysts into unfiltered water supplies has previously been implicated in the cause of several toxoplasmosis outbreaks in humans (Boyer et al. 2011; Demar et al. 2007; Vaudaux et al. 2010). But, the demonstration of definitive infection by oocysts in hosts other than humans has never been observed. The outbreak of toxoplasmosis in sea otters examined here provided a unique opportunity to study an oocyst-derived *Toxoplasma* outbreak in a natural host population. Because sea otters are listed as threatened species under the Marine Mammals Protection Act, all deaths are recorded and necropsied for cause of death determination, which provided clinical and pathological insight into the disease sequelae associated with *Toxoplasma* infection in this well-studied host. Further, as sea otters are marine mammals that do not eat warm-blooded animals, their infection was the result of oocyst contaminated water and/or bivalve food sources. The first known *Toxoplasma* infection of a southern sea otters was discovered during an outbreak caused by a novel *Toxoplasma* strain, that has been referred to as Type X, that began in the mid-late 90's (Miller et al. 2004; Kreuder et al. 2003; Cole et al. 2000).

Type X strains, originally classified as drifted Type II strains based on a single SNP at SAG1 (Cole et al. 2000), were later described by Miller et al. 2004 to be a new genotype,

referred to as Type X, because they possessed a novel, genetically distinct allele at GRA6 (Miller et al. 2004), suggesting these strains were recombinants from a cross between a Type II strain, and something new, that was later referred to as the "γ" lineage (Grigg and Sundar 2009). Type X was then sub-divided into Type X and Type A by a single SNP present at the GRA6 locus (Sundar et al. 2008). Type X was further re-classified as the fourth clonal lineage in North America (also known as HG12) based primarily on polymorphisms identified at GRA6 and GRA7 that distinguished Type X from Type II (Sundar et al. 2008; Khan, Dubey, et al. 2011).

Type X are widespread globally, although work done here focuses predominantly on Type X isolates collected only from the western coast of the United States (Dubey et al. 2014; Dubey et al. 2011; Khan, Dubey, et al. 2011; Miller et al. 2004; Miller et al. 2008; Parameswaran et al. 2010; Sundar et al. 2008; VanWormer et al. 2014). The primary aim of this study, reported in Chapter 3, was to determine whether the fourth clonal lineage is likewise expanding largely sexually by some combination of unisexual or self-mating, as was observed for Type II (Dubey et al. 2011; Khan, Dubey, et al. 2011; Miller et al. 2004; Miller et al. 2008; Sundar et al. 2008).

PCR-DNA sequencing of 53 sea otter isolates at 20 linked and unlinked markers distributed throughout the genome established that for any given marker, either a Type II allele or one of two unique allelic types existed. Phylogenetic analyses performed at each of the typing markers indicated a total lack of strain-specific SNPs, suggesting insufficient time had elapsed to generate novel SNPs *de novo* by genetic drift derived from polymerase errors during mitotic replication. The designation of two unique allelic types was supported by bootstrap analyses performed on the phylogenetic trees and strains were classified as having either γ or δ ancestry alleles following the convention used for previously characterized α and β ancestry in the Type I and III clonal lineages (Boyle et al. 2006). Importantly, across both linked and unlinked markers,

incongruence among the tree topologies was observed, with at least 10 distinct haplotypes resolved within the Type X clade, parsimonious with the isolates existing as a recombinant clade of strains that arose either by a cross between a Type II strain with a mosaic parent of both γ and δ ancestry, or from two independent crosses between a Type II strain and strains bearing either γ or δ ancestry. The absence of genetic drift at each marker, and the recent emergence of the Type X clade of strains in sea otters is more consistent with a single natural cross with a mosaic γ/δ parent, as the strains resembled F1 progeny from one or a small number of recent limited crosses.

To examine this possibility at finer resolution, whole genome sequencing was done on 21 of the Type X strains that grew well in culture, with each of the 10 distinct haplotypes being represented by at least one strain. WGS analysis identified only a limited number of crossover points in support of recent recombination between a Type II parent and at least one other parent of mixed ancestry (Khan et al. 2014; Shaik et al. 2015). Hence, the data was consistent with Type X resembling a reticulated clade of strains that are related progeny from at least one sexual cross. However, the majority of haplotypes appeared to share similar crossover points, which could not be readily explained by the presence of recombination hotspots within the genome, as these are not predicted to exist within *Toxoplasma* (Khan et al. 2014). Limited haploblock crossing over combined with low levels of genetic drift are consistent with Type X strains being derived from a recent sexual event. This sexual event was likely followed by a limited period of either asexual replication or uniparental mating as relatively few strain-specific SNPs within the Type X strains were identified genome-wide prior to the sea otter outbreak.

Strains isolated from the otters possessed a wide-range of disease sequelae, of which toxoplasmic encephalitis was described for only a subset of seropositive otters. To determine whether parasite haplotype correlated with disease, all sequenced strains were assessed for their

virulence phenotype in mice. Otter derived strains that possessed Type II genotypes were avirulent in mice, as expected. In contrast, only a subset of Type II by γ/δ recombinant strains were highly pathogenic to mice, and this pathogenicity segregated by haplotype. Importantly, the haplotype that had expanded in otters, comprising nearly 50% of otter infections (the majority of which had stranded for reasons other than protozoal encephalitis) was highly virulent in mice, suggesting that a host-shift adaptation had occurred.

Type X Contains Novel Virulence Gene Alleles

Type X strains examined in this thesis displayed three different virulence phenotypes in mice: avirulent, intermediate virulent (in which approximately half of infected mice died of acute infection, and half survived), or highly virulent (in which all mice died within 15 days of infection). Virulence shifts among cross progeny, such as the natural cross described herein, have been observed previously to occur during sexual crosses (Saeij et al. 2006; Saeij et al. 2007; Taylor et al. 2006). Forward genetic experiments to associate parasite genes responsible for causing virulence shifts have identified a suite of parasite-derived, secreted polymorphic effector proteins that hijack host immune signaling pathways to favor parasite pathogenesis (Boothroyd 2013; Dubremetz and Lebrun 2012). The majority of these previously characterized virulence phenotypes have been attributed to the inheritance of certain alleles from two proteins, ROP5 and ROP18, that in combination promote a virulence-enhancing effect in murine hosts (Behnke, Dubey, and Sibley 2016; Behnke et al. 2015). However, the allele combinations found in the Type X strains examined in this thesis, identified only avirulent allelic combinations suggesting that other virulence factors were responsible for regulating Type X murine pathogenesis. Because Type X strains resembled F1 progeny from a natural cross between two parents, they

were utilized to investigate whether it was possible to take a forward genetic approach using natural isolates of *Toxoplasma* to identify the Type X encoded murine virulence factors, or genes that resulted in the epistatic regulation of murine virulence factors. Additionally, because the Type X haplotype that was mouse-virulent had expanded as a chronic, largely benign infection in sea otter hosts, this work provided a rationale for how mouse virulence genes can be maintained cryptically in the parasite population, by allowing certain pathogenic strains of one animal host to expand as natural, chronic infections in another intermediate host.

The majority of *Toxoplasma* virulence genes previously identified using forward genetic approaches utilized quantitative trait loci (QTL) analysis of F1 progeny from sexual crosses with known parental strains. Although one parent is known for the murine virulent Type X strain, the other parent can only be inferred in aggregate, by reconstructing a mosaic genome for it from among all Type X strains. Regardless, we tested whether the isolates could be screened as if they were F1 progeny from an experimental cross to identify candidate virulence genes via QTL analysis. We anticipated that the existence of strain-specific SNPs within each Type X isolate, although relatively few in number, had the potential to confound the confidence of finding a specific genomic region by recombination that was associated with murine virulence (Smith et al. 2009). As expected, the QTL analysis of Type X strains produced many more virulence-associated genomic regions than would be expected for true F1 cross progeny, but each region existed as a discrete peak, with very little background between each peak (Behnke et al. 2015; Behnke, Khan, and Sibley 2015; Shastri et al. 2014). When top candidate genes from each genomic region were filtered based on their open reading frame, signal peptide prediction, and transmembrane domain, a manageable number of genes (*i.e.,* <50) were identified to employ reverse genetic testing to determine whether the predicted gene was associated with murine

virulence. The top candidate identified was ROP33, selected based on previously predicted activity that it could influence murine virulence *in vivo* (Peixoto et al. 2010; Wang, Li, et al. 2017). Targeted deletion of the ROP33 gene in a Type I strain engineered to be deficient in ROP18 expression established that ROP33 is a virulence locus. RH is a mouse virulent strain, and deletion of ROP33 resulted in significant attenuation of its virulence kinetics, with the majority of mice surviving acute infection, whereas mice infected with the parental RH strain all succumbed to acute infection by Day 15.

Importantly, treating the Type X isolates as F1 progeny from a cross indicated that natural isolates can be utilized for QTL approaches to identify novel virulence alleles. Future work will undoubtedly test for additional candidate virulence-associated genes in the other genomic regions with high LOD scores that were identified by the QTL. Genome-wide association studies (GWAS) could likewise be done to narrow the list of putative virulence genes and it has the added benefit of statistically accounting for both meiotic and mitotic variation between strains. However, GWAS analysis traditionally requires a larger number of whole genome sequenced isolates than currently available within the Type X clade to increase the chance of finding significantly associated genes, so it is not currently a feasible to utilized GWAS with only the Type X strains to find additional virulence genes. (Pearson and Manolio 2008; Penman et al. 2010). A more straight-forward approach to narrow the predicted number of genomic regions encoding virulence genes would be to perform experimental crosses between Type II and Type X strains to generate true F1 cross progeny. However, as discussed in Chapter 4, too few recombinant progeny were recovered from the four crosses done to perform an accurate QTL analysis.

Given the success with performing a QTL using the naturally-derived population of *T. gondii* strains infecting sea otters to identify mouse virulence genes, it is conceivable to likewise identify virulence genes associated with sea otter virulence using the same approach. Identification of mustelid virulence factors could potentially identify new approaches to combat protozoal infection and facilitate the conservation of this threatened species. Additionally, as mortality of the critically endangered Hawaiian monk seal has also been causally linked to *Toxoplasma* infection, a vaccine could dramatically reduce mortality within other endangered species threatened by protozoal infection. Especially since these parasites are a major cause of reproductive loss (Barbieri et al. 2016; Honnold et al. 2005). However, there are several significant obstacles to the characterization of mustelid virulence genes. Whilst mustelid tissue culture could allow *in vitro* analysis of novel virulence alleles for this endangered host, it is not currently available. Furthermore, inactivating the gene product to make an attenuated strain, or using the gene, to vaccinate marine mammals is problematic. Only a single vaccine is currently available against this parasite and it is not without risk. Although sheep are readily vaccinated against *Toxoplasma,* some sheep die from the vaccine, so the mortality risk of the vaccine is not acceptable for application to an endangered species such as the southern sea otter (Zhang et al. 2015; Verma and Khanna 2014). Thus, any derived virulence genes are unlikely to be able to be tested for efficacy within the intended mustelid hosts.

Biased Mating May Explain the Distinct Clonality Apparent Globally

Due to methods used previously to hatch sporulated oocysts into sporozoites and select immediately using double drug regimens, it was impossible to determine the true recombination rate for *Toxoplasma* sexual crosses. This is because all previous studies isolated individual

progeny from pooled sporozoites that had been selectively expanded under double drug selection prior to limiting dilution to select for individual recombinant progeny. A new, macrophage-facilitated method of excystation was developed here to increase yield and independently isolate progeny from a sexual cross. These progeny were isolated without drug-selection or competitive growth. This novel procedure produced increased numbers of individually isolated progeny and reduced the hazards to researchers associated with excystation. Further, novel WGS of individually isolated progeny increased the depth of genomic resolution available and allowed for the unbiased determination of recombination rates within *Toxoplasma*.

Based upon whole genome sequence analysis of the progeny from four independent crosses, parental recombination rates were determined to have occurred at a significantly lower frequency than previously reported (Khan et al. 2014; Khan et al. 2005; Pfefferkorn and Pfefferkorn 1980; Saeij et al. 2007; Pfefferkorn, Pfefferkorn, and Colby 1977). This work is the first to examine progeny from a low-recombination frequency cross as all other crosses published displayed either a larger number of outcrossed progeny, or suggested that some strains, which failed to outcross and only produced detectable numbers of oocysts by self-mating, are mating incompetent. WGS of all individually isolated progeny allowed self-mating and outcrossing to be interrogated from a single experimental sexual cross and facilitated the determination that *Toxoplasma* is more likely to self-mate than outcross, even in co-infection.

A bias in isolation techniques could skew the isolated progeny toward one parental strain, but this is unlikely for the isolations done here as excystations were done in parallel for the four independent crosses examined, and while the recombination frequencies were similar, the parental strain that was preferentially expanded by self-mating was different for each cross, and no skewing to one parental type was identified. The greatest risk for bias from previous studies

was due to the pooled competitive growth of strains under double drug selection during cellular culture prior to isolation, which was removed in the new excystation methods detailed in this work. These improved individual isolation techniques abrogated growth competition by separating progeny from one another before growth in culture. However, progeny must still grow in host cells, which could bias toward strains better suited to expand in cell culture (Boyle and Radke 2009).

Growth in the feline gut prior to sexual replication could also alter recombination. The contribution of the feline microbiome to these crosses is unknown, and bacterial factors have been shown to be important in regulating gametocytogenesis in other eukaryotes, such as choanoflagellates (Woznica et al. 2016; Woznica et al. 2017). In *Eimeria* spp., strains that colonize the same region of the GI tract tend to cross with each other, and this phenomenon is the basis for speciation among the chicken coccidia (Walker et al. 2013). If different strains of *Toxoplasma* colonize different regions of the cat lumen in a strain-specific manner, this may impact outcrossing, and favor inbreeding, or self-mating, as more diverse strains are not in close proximity physically or temporally to fuse to produce oocysts. Macrogamete production could also contribute to a bias against outcrossing as there are not enough microgametes to ensure cross-fertilization within any given sexual cross (Ferguson 2002; Ferguson et al. 1974; Ferguson, Hutchison, and Siim 1975; West, Smith, and Read 2000). Mating-type loci have not been formally demonstrated to exist in *Toxoplasma*, but in fungi, they are important variables that determine sexual competency and mating compatibility (Feretzaki and Heitman 2013; Heitman 2010; Ni et al. 2011; Sun et al. 2014).

It is likely that biased recombination is more prevalent in *Toxoplasma* than previously predicted. In fact, the uniparental-biased recombination rate of 2-3% recombination determined

here for *Toxoplasma* is similar to recombination rates known from other protozoan parasites which possess non-Mendelian recombination frequencies. In *Leishmania* recombination rates can be as low as 6% where it has been hypothesized that biased recombination rates are highly dependent on the species of the sand-fly definitive host the parasite infects (Inbar et al. 2013; Romano et al. 2014). Based on the similarity of the genetic hybridization rates between these two parasites, it is possible that similar mechanisms are at play within the definitive felid host of *Toxoplasma*, which possesses a highly diverse genus, where these may influence the successful hybridization of gametes.

Elucidation of the extent to which *Toxoplasma* undergoes outcrossing in nature or is expanding clonally by asexual division versus unisexual expansion by self-mating within a cat is important data to not only initiate transmission blocking interventions, but also shed light on the origin of the clonal population structure for *Toxoplasma* in Europe and North America. The geographic origin of clonality, especially between North and South America, is still under debate. Some studies indicate that North American strains were seeded by South American parasite migration (Bertranpetit et al. 2016; Walzer and Boyle 2012), while other studies determined that clonal strains originated in North America (established by ancestral migration from Europe) (Fux et al. 2007; Sibley et al. 2009). These models and all previous mutation rates, however, were predicted based on the known *Plasmodium* mutation rates at that time, which have since been shown to be incorrect (Rich et al. 1998; Joy et al. 2003). Furthermore, these earlier calculations are inherently inaccurate for *Toxoplasma* due to lifecycle differences between the two parasites. *Plasmodium* has a mandatory sexual replication cycle for propagation, whereas *Toxoplasma* can bypass its sexual cycle and remain dormant as bradyzoite cysts for extended periods of time and can expand between intermediate hosts asexually by carnivory. Further, self-

mating in the definitive host phenocopies the asexual expansion model and causes difficulties in differentiating asexual and unisexual growth (Wendte, Miller, Lambourn, et al. 2010; Boyer et al. 2011; Jones and Dubey 2010). As evidenced in Chapter 4, the bias for uniparental expansion in the crossing experiments, as well as the unisexual mating observed between Type II strains, would produce a mainly clonal population structure, that would be indistinguishable from a parasite that expands exclusively asexually (Ferguson et al. 1974; West, Smith, and Read 2000; Grigg and Sundar 2009).

It should be possible to calculate a meiotic replication mutation rate using the uniparental progeny isolated in Chapter 4. Meiotic mutation rates have not previously been calculated for *Toxoplasma but* are hypothesized to contribute to the maintenance of the apparently clonal population structure (Minot et al. 2012; Roach and Heitman 2014; Wendte, Miller, Lambourn, et al. 2010). Experiments on the mitotic mutation rate are currently in progress using *in vitro* serial passage of a self-mated strain based on the more precise experimental methodologies developed in *Plasmodium* (Bopp et al. 2013). Together, mutation rates for meiotic sexual self-mating and mitotic asexual reproduction should enable a more accurate mathematical model for the genomic rate of evolution for *Toxoplasma* (Bopp et al. 2013; Claessens et al. 2014).

Toxoplasma has a Diverse Population Structure Prone to Biased and Unisexual Recombination, as well as Mitotic Drift

Clonality has been observed within natural parasite populations. However, the frequency and effect of sexual recombination in shaping these clonal populations remains to be fully interrogated by WGS. Several protozoan parasites possess the genetic capacity to undergo meiotic recombination yet for unknown reasons their natural genetic population structure that is

predominantly clonal, such as *Cryptococcus* (Billmyre et al. 2014), *Cryptosporidium* (Awad-El-Kariem 1999), and *Neisseria* (Budroni et al. 2011). Drug-resistance has been shown to select for clonality in *Plasmodium* (Ariey et al. 2014; Miotto et al. 2013; Walliker 1991). However, it is unknown if these populations are maintained once drug-selection is removed.

In *Leishmania,* the sheer number of vector species that serve as definitive hosts for the variety of *Leishmania* strains and species may limit interbreeding, especially because many vector species do not share overlapping host ranges which could pose a significant barrier for mixing to occur and thereby promote the maintenance of clonal lineages. When outcrossing does occur, other mechanisms could also bias toward a clonal population structure such as copy number variation and aneuploidy (Iantorno et al. 2017) which have been proposed to purify genomic heterozygosity to actively facilitate genomic clonality (Inbar et al. 2013; Romano et al. 2014; Ramirez and Llewellyn 2014). Felid diversity has likewise been proposed to influence what strains of *Toxoplasma* can be expanded in any one geographic region, which may contribute to clonal population genetic structures depending on the felid species present (Khan et al. 2006; Fux et al. 2007; Khan, Miller, et al. 2011).

All previous population studies in *Toxoplasma* have focused on the genetic diversity identified across the population and not within a clonal lineage (Lorenzi et al. 2016). However, *Toxoplasma* clades are more genetically diverse within a clade than previously expected, as evidenced by the genomic diversity identified within both the Type II and X clades shown here. The elevated genetic diversity identified among Type II clones established that mitotic drift has positively influenced within-clade genetic diversity among circulating strains of *Toxoplasm*a. Additionally, the intra-clade specific sexual recombination detected in some North American *Toxoplasma* isolates, that appeared to be mosaic admixtures with strains circulating in Europe or

Africa, identified a previously underappreciated rate of genetic hybridization within clonal lineages. Further, the inheritance of apicoplast sequences that were of European or African decent within these admixture lines isolated from North American animals supported this conclusion (Burrells et al. 2013; Grigg and Sundar 2009; Howe and Sibley 1995; Khan, Dubey, et al. 2011; Fux et al. 2007; Minot et al. 2012; Su et al. 2012).

Sexual crossing has been shown to produce novel combinations of virulence genes that can alter the parasite's pathogenesis and allow infection to occur in novel hosts (Adomako-Ankomah et al. 2014; Boothroyd 2009; Grigg and Sundar 2009; Su et al. 2003; Grigg, Bonnefoy, et al. 2001; Fernandez et al. 2016; Hakimi, Olias, and Sibley 2017). It is likely that sexual recombination is the predominant mechanism used by generalist parasites to adapt to and infect new host species. Bringing virulence alleles into new combinations may allow for effective expansion of the parasite's host range, facilitating host-shifts for generalist parasites like *Toxoplasma*. These host-shifts may explain how a mouse virulent *Toxoplasma* strain can expand cryptically in one animal host, such as a sea otter, yet possess just the right combination of virulence alleles to inactivate murine host immunity and cause acute epidemic disease. The host-shift identified among the Type X progeny that expanded in sea otters is not an anomalous finding, as it has likewise been demonstrated that other murine virulent strains, such as Type I, which have selectively expanded as an avirulent infection in a variety of avian hosts (Miller, Grigg unpublished).

The conclusions presented herein are based on whole genome sequencing, rather than the more traditional PCR-RFLP marker characterization. To determine if the diversity observed in the Type II and X clades is extant across the *Toxoplasma* population, WGS of additional isolates from other clonal lineages will be necessary (Pan et al. 2012; Parameswaran et al. 2010; Yang et

al. 2013; Ajzenberg et al. 2010). Ideally, strains would be sequenced and genomes *de novo* assembled. Although *de novo* assembly provides more accurate genomic comparisons, it requires both longer sequencing reads and higher read depth, making this technique prohibitively expensive for multiple strains. Comparison of *Toxoplasma* strains by reference mapping is not without its own inherent biases and these limitations must be considered when evaluating datasets. For example, as with many *de novo* assembled genomes, the common *Toxoplasma* reference strain (ME49) does not accurately annotate low-diversity tandemly arrayed genes, often stacking them rather than correctly assembling them as tandem repeats end-to-end (Matrajt et al. 1999; Adomako-Ankomah et al. 2014). These regions must be interpreted with caution. Additionally, genomic regions with known sequence gaps can lead to erroneous mapping of sequencing reads. Finally, genes with strain-specific indels longer than the sequencing reads, such as GRA15, are often incorrectly assembled during reference mapping, and can lead to an elevated SNP diversity that is not actually present between genomes (Rosowski et al. 2011). Despite these limitations, evaluating single nucleotide polymorphisms at the level of chromosomes lends computational clarity and evolutionary resolution to WGS strains beyond that provided by PCR-RFLP. Furthermore, aligning to a shared (Type II) reference biases all non-Type II strains equally. The benefits of increased genomic depth for characterization far out-weigh the potential pitfalls of WGS characterization.

Further Work on Toxoplasma Sexual Cycles is Necessary to Elucidate the Lifecycle Diversity and Cryptic Virulence of this Sexual Parasite

Toxoplasma has a uniquely flexible lifecycle whereby it can undergo either sexual replication or asexual expansion in any order or combination. With the increased appreciation

that sexual recombination is occurring cryptically either via unisexual mating or with a bias toward self-mating when the parasite has the opportunity to out-cross, identifying an accurate mathematical model for the evolution of *Toxoplasma* is important to limit transmission and examine the evolution of new parasite lines. Evidence herein has shown that *Toxoplasma* is not expanding strictly asexually within clonal clades such as Type II and X. The fact that this parasite, when given the chance to out-cross, predominantly self-mates argues that barriers must exist that limit productive recombination. Whether mating types exist is certainly a possibility that has not been definitively disproven. But equally, sexual bias could be related to physical barriers that limit the ability of gametes from two different strains to find each other within the definitive host gut lumen. While it remains to be tested whether mating preferences exist between distinct clades of *Toxoplasma,* this phenomenon would explain how sexual expansion of clones seem to preferentially self-mate given the opportunity to outcross. How *Toxoplasma* parasites chose their partners for sexual recombination is currently unknown, in large part due to the black box that is the feline gut where gametocytogenesis and mating occurs. In yeast, sexual partner selection is determined by genomic mating types (Ni et al. 2011). However, *Toxoplasma* displays no evidence for mating types within their genomes. Other parasites, as well as a number of fungal species maintain similarly cryptic sexual replication cycles like *Toxoplasma.* It is possible that work examining the replication cycles of these fungi may elucidate conserved mechanisms for unisexual (or similar strain) mating that could provide perspective for the biased nature of sexual replication in *Toxoplasma* (Heitman 2010; Roach and Heitman 2014).

As *Toxoplasma* infections have previously been shown to differentially respond to stress in the cellular environment, it is possible that *Toxoplasma* can switch between outcrossing and unisexual or self-mating based on an unknown signal within the feline enterocytes (Hakimi,

Olias, and Sibley 2017; Yang et al. 2015). Experiments to elucidate the activity of merozoites and derived oocysts in the feline definitive host are necessary to elucidate the host environmental signals involved in the biased production of *Toxoplasma* progeny. Although the majority of strains in the *Toxoplasma* population are derived from the known "clonal" lineages, the more diverse strains are often the most problematic, giving rise to disease outbreaks following sexual crossing (Vaudaux et al. 2010). Thus, it is imperative that an improved understanding of the population dynamics of *Toxoplasma* be determined to better understand these divergent strains.

Likely, these divergent strains of *Toxoplasma* contain novel virulence genes. Hence, sexual progeny between divergent and "clonal" strains are necessary to further elucidate novel virulence factors within the *Toxoplasma* population. Furthermore, because it is hypothesized that divergent strains are best suited for different hosts than the clonal clades, testing these progeny against an array of intermediate host species could elucidate virulence genes important to host-shifts. A genetic mechanism favoring one host over another would verify that intermediate host selection is a driving factor for the evolution of this generalist parasite (Boothroyd 2009). Potential host-shift genes need to be elucidated and their functions characterized to understand how this generalist pathogen maintains its cryptic virulence to maintain the ability to invade and manipulate the host environment across *Toxoplasma's* global range of hosts.

Bibliography

Adomako-Ankomah, Y., E. D. English, and JJ Danielson. 2016. 'Host Mitochondrial Association Evolved in the Human Parasite Toxoplasma gondii via Neofunctionalization of a Gene Duplicate', *Genetics*.

Adomako-Ankomah, Y., G. M. Wier, A. L. Borges, H. E. Wand, and J. P. Boyle. 2014. 'Differential locus expansion distinguishes Toxoplasmatinae species and closely related strains of Toxoplasma gondii', *MBio*, 5: e01003-13.

Agrawal, A. F. 2006. 'Evolution of sex: why do organisms shuffle their genotypes?', *Curr Biol*, 16: R696-704.

Ajzenberg, D., A. L. Banuls, M. Tibayrenc, and M. L. Darde. 2002. 'Microsatellite analysis of Toxoplasma gondii shows considerable polymorphism structured into two main clonal groups', *Int J Parasitol*, 32: 27-38.

Ajzenberg, D., F. Collinet, A. Mercier, P. Vignoles, and M. L. Darde. 2010. 'Genotyping of Toxoplasma gondii isolates with 15 microsatellite markers in a single multiplex PCR assay', *J Clin Microbiol*, 48: 4641-5.

Ajzenberg, D., H. Yera, P. Marty, L. Paris, F. Dalle, J. Menotti, D. Aubert, J. Franck, M. H. Bessieres, D. Quinio, H. Pelloux, L. Delhaes, N. Desbois, P. Thulliez, F. Robert-Gangneux, C. Kauffmann-Lacroix, S. Pujol, M. Rabodonirina, M. E. Bougnoux, B. Cuisenier, C. Duhamel, T. H. Duong, D. Filisetti, P. Flori, F. Gay-Andrieu, F. Pratlong, G. Nevez, A. Totet, B. Carme, H. Bonnabau, M. L. Darde, and I. Villena. 2009. 'Genotype of 88 Toxoplasma gondii isolates associated with toxoplasmosis in immunocompromised patients and correlation with clinical findings', *J Infect Dis*, 199: 1155-67.

Alaganan, A., S. J. Fentress, K. Tang, Q. Wang, and L. D. Sibley. 2014. 'Toxoplasma GRA7 effector increases turnover of immunity-related GTPases and contributes to acute virulence in the mouse', *Proc Natl Acad Sci U S A*, 111: 1126-31.

Ariey, F., B. Witkowski, C. Amaratunga, J. Beghain, A. C. Langlois, N. Khim, S. Kim, V. Duru, C. Bouchier, L. Ma, P. Lim, R. Leang, S. Duong, S. Sreng, S. Suon, C. M. Chuor, D. M. Bout, S. Menard, W. O. Rogers, B. Genton, T. Fandeur, O. Miotto, P. Ringwald, J. Le Bras, A. Berry, J. C. Barale, R. M. Fairhurst, F. Benoit-Vical, O. Mercereau-Puijalon, and D. Menard. 2014. 'A molecular marker of artemisinin-resistant Plasmodium falciparum malaria', *Nature*, 505: 50-5.

Arisue, N., and T. Hashimoto. 2014. 'Phylogeny and evolution of apicoplasts and apicomplexan parasites', *Parasitol Int*.

Avelino, M. M., W. N. Amaral, I. M. Rodrigues, A. R. Rassi, M. B. Gomes, T. L. Costa, and A. M. Castro. 2014. 'Congenital toxoplasmosis and prenatal care state programs', *BMC Infect Dis*, 14: 33.

Awad-El-Kariem, F. M. 1999. 'Does Cryptosporidium parvum have a Clonal Population Structure?', *Parasitology Today*, 15.

Ayi, I., K. D. Kwofie, E. A. Blay, J. H. Osei, K. K. Frempong, R. Koku, A. Ghansah, M. Lartey, T. Suzuki, D. A. Boakye, K. A. Koram, and N. Ohta. 2016. 'Clonal types of Toxoplasma gondii among immune compromised and immune competent individuals in Accra, Ghana', *Parasitol Int*, 65: 238-44.

Barbieri, M. M., L. Kashinsky, D. S. Rotstein, K. M. Colegrove, K. H. Haman, S. L. Magargal, A. R. Sweeny, A. C. Kaufman, M. E. Grigg, and C. L. Littnan. 2016. 'Protozoal-related mortalities in endangered Hawaiian monk seals Neomonachus schauinslandi', *Dis Aquat Organ*, 121: 85-95.

Barbosa, L., C. K. Johnson, D. M. Lambourn, A. K. Gibson, K. H. Haman, J. L. Huggins, A. R. Sweeny, N. Sundar, S. A. Raverty, and M. E. Grigg. 2015. 'A novel Sarcocystis neurona genotype XIII is associated with severe encephalitis in an unexpectedly broad range of marine mammals from the northeastern Pacific Ocean', *Int J Parasitol*.

Behnke, M. S., J. P. Dubey, and L. D. Sibley. 2016. 'Genetic Mapping of Pathogenesis Determinants in Toxoplasma gondii', *Annu Rev Microbiol*, 70: 63-81.

Behnke, M. S., S. J. Fentress, M. Mashayekhi, L. X. Li, G. A. Taylor, and L. D. Sibley. 2012. 'The Polymorphic Pseudokinase ROP5 Controls Virulence in Toxoplasma gondii by Regulating the Active Kinase ROP18', *PLoS Pathog*, 8: e1002992.

Behnke, M. S., A. Khan, E. J. Lauron, J. R. Jimah, Q. Wang, N. H. Tolia, and L. D. Sibley. 2015. 'Rhoptry Proteins ROP5 and ROP18 Are Major Murine Virulence Factors in Genetically Divergent South American Strains of Toxoplasma gondii', *PLoS Genet*, 11: e1005434.

Behnke, M. S., A. Khan, and L. D. Sibley. 2015. 'Genetic Mapping Reveals that Sinefungin Resistance in Toxoplasma gondii Is Controlled by a Putative Amino Acid Transporter Locus That Can Be Used as a Negative Selectable Marker', *Eukaryot Cell*, 14: 140-8.

Behnke, M. S., A. Khan, J. C. Wootton, J. P. Dubey, K. Tang, and L. D. Sibley. 2011. 'Virulence differences in Toxoplasma mediated by amplification of a family of polymorphic pseudokinases', *Proc Natl Acad Sci U S A*, 108: 9631-6.

Bellanger, X., S. Payot, N. Leblond-Bourget, and G. Guedon. 2014. 'Conjugative and mobilizable genomic islands in bacteria: evolution and diversity', *FEMS Microbiol Rev*, 38: 720-60.

Bertranpetit, E., T. Jombart, E. Paradis, H. Pena, J. Dubey, C. Su, A. Mercier, S. Devillard, and D. Ajzenberg. 2016. 'Phylogeography of Toxoplasma gondii points to a South American origin', *Infect Genet Evol*, 48: 150-55.

Billmyre, R. B., D. Croll, W. Li, P. Mieczkowski, D. A. Carter, C. A. Cuomo, J. W. Kronstad, and J. Heitman. 2014. 'Highly recombinant VGII Cryptococcus gattii population develops clonal outbreak clusters through both sexual macroevolution and asexual microevolution', *MBio*, 5: e01494-14.

Blackston, C. R., J. P. Dubey, E. Dotson, C. Su, P. Thulliez, D. Sibley, and T. Lehmann. 2001. 'High-Resolution Typing of Toxoplasma gondii Using Microsatellite Loci', *Journal of Parasitology*, 87: 1472-75.

Bohne, W., J. Heesemann, and U. Gross. 1994. 'Reduced Replication of Toxoplasma-Gondii Is Necessary for Induction of Bradyzoite-Specific Antigens - a Possible Role for Nitric-Oxide in Triggering Stage Conversion', *Infection and Immunity*, 62: 1761-67.

Boothroyd, J. C. 2009. 'Expansion of host range as a driving force in the evolution of Toxoplasma', *Mem Inst Oswaldo Cruz*, 104: 179-84.

———. 2013. 'Have it your way: how polymorphic, injected kinases and pseudokinases enable Toxoplasma to subvert host defenses', *PLoS Pathog*, 9: e1003296.

Boothroyd, J. C., and J. F. Dubremetz. 2008. 'Kiss and spit: the dual roles of Toxoplasma rhoptries', *Nat Rev Microbiol*, 6: 79-88.

Boothroyd, J. C., and M. E. Grigg. 2002. 'Population biology of Toxoplasma gondii and its relevance to human infection: do different strains cause different disease?' in, *Curr Opin Microbiol* (England).

Bopp, S. E., M. J. Manary, A. T. Bright, G. L. Johnston, N. V. Dharia, F. L. Luna, S. McCormack, D. Plouffe, C. W. McNamara, J. R. Walker, D. A. Fidock, E. L. Denchi, and E. A. Winzeler. 2013. 'Mitotic evolution of Plasmodium falciparum shows a stable core genome but recombination in antigen families', *PLoS Genet*, 9: e1003293.

Boyer, K., D. Hill, E. Mui, K. Wroblewski, T. Karrison, J. P. Dubey, M. Sautter, A. G. Noble, S. Withers, C. Swisher, P. Heydemann, T. Hosten, J. Babiarz, D. Lee, P. Meier, R. McLeod, and Group Toxoplasmosis Study. 2011. 'Unrecognized ingestion of Toxoplasma gondii oocysts leads to congenital toxoplasmosis and causes epidemics in North America', *Clin Infect Dis*, 53: 1081-9.

Boyle, J. P., B. Rajasekar, J. P. Saeij, J. W. Ajioka, M. Berriman, I. Paulsen, D. S. Roos, L. D. Sibley, M. W. White, and J. C. Boothroyd. 2006. 'Just one cross appears capable of dramatically altering the population biology of a eukaryotic pathogen like Toxoplasma gondii', *Proc Natl Acad Sci U S A*, 103: 10514-9.

Boyle, Jon P., and Jay R. Radke. 2009. 'A history of studies that examine the interactions of Toxoplasma with its host cell: Emphasis on in vitro models', *Int J Parasitol*, 39: 903-14.

Brabin, B. J. 1985. 'Epidemiology of infection in pregnancy', *Rev Infect Dis*, 7: 579-603.

Braun, L., M. P. Brenier-Pinchart, M. Yogavel, A. Curt-Varesano, R. L. Curt-Bertini, T. Hussain, S. Kieffer-Jaquinod, Y. Coute, H. Pelloux, I. Tardieux, A. Sharma, H. Belrhali, A. Bougdour, and M. A. Hakimi. 2013. 'A Toxoplasma dense granule protein, GRA24, modulates the early immune response to infection by promoting a direct and sustained host p38 MAPK activation', *Journal of Experimental Medicine*, 210: 2071-86.

Budroni, S., E. Siena, J. C. D. Hotopp, K. L. Seib, D. Serruto, C. Nofroni, M. Comanducci, D. R. Riley, S. C. Daugherty, S. V. Angiuoli, A. Covacci, M. Pizza, R. Rappuoli, E. R. Moxon, H. Tettelin, and D. Medini. 2011. 'Neisseria meningitidis is structured in clades associated with restriction modification systems that modulate homologous recombination', *Proc Natl Acad Sci U S A*, 108: 4494-99.

Burrells, A., P. M. Bartley, I. A. Zimmer, S. Roy, A. C. Kitchener, A. Meredith, S. E. Wright, E. A. Innes, and F. Katzer. 2013. 'Evidence of the three main clonal Toxoplasma gondii lineages from wild mammalian carnivores in the UK', *Parasitology*, 140: 1768-76.

Butcher, B. A., B. A. Fox, L. M. Rommereim, S. G. Kim, K. J. Maurer, F. Yarovinsky, D. R. Herbert, D. J. Bzik, and E. Y. Denkers. 2011. 'Toxoplasma gondii rhoptry kinase ROP16 activates STAT3 and STAT6 resulting in cytokine inhibition and arginase-1-dependent growth control', *PLoS Pathog*, 7: e1002236.

Caffaro, C. E., and J. C. Boothroyd. 2011. 'Evidence for host cells as the major contributor of lipids in the intravacuolar network of Toxoplasma-infected cells', *Eukaryot Cell*, 10: 1095-9.

Carruthers, V. B., and L. D. Sibley. 1997. 'Sequential protein secretion from three distinct organelles of Toxoplasma gondii accompanies invasion of human fibroblasts', *Eur J Cell Biol*, 73: 114-23.

Cavailles, P., P. Flori, O. Papapietro, C. Bisanz, D. Lagrange, L. Pilloux, C. Massera, S. Cristinelli, D. Jublot, O. Bastien, C. Loeuillet, D. Aldebert, B. Touquet, G. J. Fournie, and M. F. Cesbron-Delauw. 2014. 'A highly conserved Toxo1 haplotype directs

resistance to toxoplasmosis and its associated caspase-1 dependent killing of parasite and host macrophage', *PLoS Pathog*, 10: e1004005.

Cavailles, P., V. Sergent, C. Bisanz, O. Papapietro, C. Colacios, M. Mas, J. F. Subra, D. Lagrange, M. Calise, S. Appolinaire, T. Faraut, P. Druet, A. Saoudi, M. H. Bessieres, B. Pipy, M. F. Cesbron-Delauw, and G. J. Fournie. 2006. 'The rat Toxo1 locus directs toxoplasmosis outcome and controls parasite proliferation and spreading by macrophage-dependent mechanisms', *Proc Natl Acad Sci U S A*, 103: 744-49.

Cirelli, K. M., G. Gorfu, M. A. Hassan, M. Printz, D. Crown, S. H. Leppla, M. E. Grigg, J. P. Saeij, and M. Moayeri. 2014. 'Inflammasome sensor NLRP1 controls rat macrophage susceptibility to Toxoplasma gondii', *PLoS Pathog*, 10: e1003927.

Claessens, A., W. L. Hamilton, M. Kekre, T. D. Otto, A. Faizullabhoy, J. C. Rayner, and D. Kwiatkowski. 2014. 'Generation of antigenic diversity in Plasmodium falciparum by structured rearrangement of Var genes during mitosis', *PLoS Genet*, 10: e1004812.

Cole, R. A., D. S. Lindsay, D. K. Howe, C. L. Roderick, J. P. Dubey, N. J. Thomas, and L. A. Baeten. 2000. 'Biological and Molecular Characterizations Oftoxoplasma Gondiistrains Obtained from Southern Sea Otters (Enhydra Lutris Nereis)', *Journal of Parasitology*, 86: 526-30.

Conrad, P. A., M. A. Miller, C. Kreuder, E. R. James, J. Mazet, H. Dabritz, D. A. Jessup, F. Gulland, and M. E. Grigg. 2005. 'Transmission of Toxoplasma: clues from the study of sea otters as sentinels of Toxoplasma gondii flow into the marine environment', *Int J Parasitol*, 35: 1155-68.

Cornelissen, A. W., and J. P. Overdulve. 1985. 'Sex determination and sex differentiation in coccidia: gametogony and oocyst production after monoclonal infection of cats with free-living and intermediate host stages of Isospora (Toxoplasma) gondii', *Parasitology*, 90 (Pt 1): 35-44.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and Group Genomes Project Analysis. 2011. 'The variant call format and VCFtools', *Bioinformatics*, 27: 2156-8.

Darde, M. L., B. Bouteille, and M. Pestre-Alexandre. 1992. 'Isoenzyme analysis of 35 Toxoplasma gondii isolates and the biological and epidemiological implications', *J Parasitol*, 78: 786-94.

Demar, M., D. Ajzenberg, D. Maubon, F. Djossou, D. Panchoe, W. Punwasi, N. Valery, C. Peneau, J. L. Daigre, C. Aznar, B. Cottrelle, L. Terzan, M. L. Darde, and B. Carme. 2007. 'Fatal outbreak of human toxoplasmosis along the Maroni River: epidemiological, clinical, and parasitological aspects', *Clin Infect Dis*, 45: e88-95.

Derouin, F., H. Pelloux, and Escmid Study Group on Clinical Parasitology. 2008. 'Prevention of toxoplasmosis in transplant patients', *Clin Microbiol Infect*, 14: 1089-101.

Dubey, J. P. 1995. 'Duration of immunity to shedding of Toxoplasma gondii oocysts by cats', *J Parasitol*, 81: 410-5.

———. 1998. 'Toxoplasma gondii oocyst survival under defined temperatures', *J Parasitol*, 84: 862-5.

———. 2001. 'Oocyst Shedding by Cats Fed Isolated Bradyzoites and Comparison of Infectivity of Bradyzoites of the VEG StrainToxoplasma gondiito Cats and Mice', *Journal of Parasitology*, 87: 215-19.

Dubey, J. P., and J. K. Frenkel. 1972. 'Cyst-induced toxoplasmosis in cats', *J Protozool*, 19: 155-77.
———. 1976. 'Feline toxoplasmosis from acutely infected mice and the development of Toxoplasma cysts', *J Protozool*, 23: 537-46.
Dubey, J. P., L. T. Huong, N. Sundar, and C. Su. 2007. 'Genetic characterization of Toxoplasma gondii isolates in dogs from Vietnam suggests their South American origin', *Vet Parasitol*, 146: 347-51.
Dubey, J. P., and J. L. Jones. 2008. 'Toxoplasma gondii infection in humans and animals in the United States', *Int J Parasitol*, 38: 1257-78.
Dubey, J. P., A. N. Patitucci, C. Su, N. Sundar, O. C. Kwok, and S. K. Shen. 2006. 'Characterization of Toxoplasma gondii isolates in free-range chickens from Chile, South America', *Vet Parasitol*, 140: 76-82.
Dubey, J. P., S. K. Shen, O. C. Kwok, and J. K. Frenkel. 1999. 'Infection and immunity with the RH strain of Toxoplasma gondii in rats and mice', *J Parasitol*, 85: 657-62.
Dubey, J. P., K. Van Why, S. K. Verma, S. Choudhary, O. C. Kwok, A. Khan, M. S. Behinke, L. D. Sibley, L. R. Ferreira, S. Oliveira, M. Weaver, R. Stewart, and C. Su. 2014. 'Genotyping Toxoplasma gondii from wildlife in Pennsylvania and identification of natural recombinants virulent to mice', *Vet Parasitol*, 200: 74-84.
Dubey, J. P., G. V. Velmurugan, C. Rajendran, M. J. Yabsley, N. J. Thomas, K. B. Beckmen, D. Sinnett, D. Ruid, J. Hart, P. A. Fair, W. E. McFee, V. Shearn-Bochsler, O. C. Kwok, L. R. Ferreira, S. Choudhary, E. B. Faria, H. Zhou, T. A. Felix, and C. Su. 2011. 'Genetic characterisation of Toxoplasma gondii in wildlife from North America revealed widespread and high prevalence of the fourth clonal type', *Int J Parasitol*, 41: 1139-47.
Dubey, J. P., X. Q. Zhu, N. Sundar, H. Zhang, O. C. Kwok, and C. Su. 2007. 'Genetic and biologic characterization of Toxoplasma gondii isolates of cats from China', *Vet Parasitol*, 145: 352-6.
Dubey, J.P. 2010. *Toxoplasmosis of Animals and Humans, Second Edition* (CRC Press).
Dubremetz, J. F., and M. Lebrun. 2012. 'Virulence factors of Toxoplasma gondii', *Microbes Infect*, 14: 1403-10.
Dunn, J. D., S. Ravindran, S. K. Kim, and J. C. Boothroyd. 2008. 'The Toxoplasma gondii dense granule protein GRA7 is phosphorylated upon invasion and forms an unexpected association with the rhoptry proteins ROP2 and ROP4', *Infection and Immunity*, 76: 5853-61.
Dupont, C. D., and C. A. Hunter. 2012. 'Guanylate-binding proteins: niche recruiters for antimicrobial effectors', *Immunity*, 37: 191-3.
Eidell, K. P., T. Burke, and M. J. Gubbels. 2010. 'Development of a screen to dissect Toxoplasma gondii egress', *Mol Biochem Parasitol*, 171: 97-103.
Elbez-Rubinstein, A., D. Ajzenberg, M. L. Darde, R. Cohen, A. Dumetre, H. Yera, E. Gondon, J. C. Janaud, and P. Thulliez. 2009. 'Congenital toxoplasmosis and reinfection during pregnancy: case report, strain characterization, experimental model of reinfection, and review', *J Infect Dis*, 199: 280-5.
English, E. D., Y. Adomako-Ankomah, and J. P. Boyle. 2015. 'Secreted effectors in Toxoplasma gondii and related species: determinants of host range and pathogenesis?', *Parasite Immunol*, 37: 127-40.

Etheridge, R. D., A. Alaganan, K. Tang, H. J. Lou, B. E. Turk, and L. D. Sibley. 2014. 'The Toxoplasma pseudokinase ROP5 forms complexes with ROP18 and ROP17 kinases that synergize to control acute virulence in mice', *Cell Host Microbe*, 15: 537-50.

Ewald, S. E., J. Chavarria-Smith, and J. C. Boothroyd. 2014. 'NLRP1 is an inflammasome sensor for Toxoplasma gondii', *Infection and Immunity*, 82: 460-8.

Fazaeli, A., P. E. Carter, M. L. Darde, and T. H. Pennington. 2000. 'Molecular typing of Toxoplasma gondii strains by GRA6 gene sequence analysis', *Int J Parasitol*, 30: 637-42.

Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt. 2004. 'eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data', *Journal of Bacteriology*, 186: 1518-30.

Fekkar, A., D. Ajzenberg, B. Bodaghi, F. Touafek, P. Le Hoang, J. Delmas, P. Y. Robert, M. L. Darde, D. Mazier, and L. Paris. 2011. 'Direct genotyping of Toxoplasma gondii in ocular fluid samples from 20 patients with ocular toxoplasmosis: predominance of type II in France', *J Clin Microbiol*, 49: 1513-7.

Feretzaki, M., and J. Heitman. 2013. 'Unisexual reproduction drives evolution of eukaryotic microbial pathogens', *PLoS Pathog*, 9: e1003674.

Ferguson, D. 2002. 'Toxoplasma gondii and sex: essential or optional extra?', *Trends Parasitol*, 18: 351.

Ferguson, D. J., F. L. Henriquez, M. J. Kirisits, S. P. Muench, S. T. Prigge, D. W. Rice, C. W. Roberts, and R. L. McLeod. 2005. 'Maternal inheritance and stage-specific variation of the apicoplast in Toxoplasma gondii during development in the intermediate and definitive host', *Eukaryot Cell*, 4: 814-26.

Ferguson, D. J., W. M. Hutchison, J. F. Dunachie, and J. C. Siim. 1974. 'Ultrastructural study of early stages of asexual multiplication and microgametogony of Toxoplasma gondii in the small intestine of the cat', *Acta Pathol Microbiol Scand B Microbiol Immunol*, 82: 167-81.

Ferguson, D. J., W. M. Hutchison, and J. C. Siim. 1975. 'The ultrastructural development of the macrogamete and formation of the oocyst wall of Toxoplasma gondii', *Acta Pathol Microbiol Scand B*, 83: 491-505.

Ferguson, David J. P., and Jean-François Dubremetz. 2014. 'Chapter 2 - The Ultrastructure of Toxoplasma gondii.' in, *Toxoplasma Gondii (Second Edition)* (Academic Press: Boston).

Fernandez, C., J. Jaimes, M. C. Ortiz, and J. D. Ramirez. 2016. 'Host and Toxoplasma gondii genetic and non-genetic factors influencing the development of ocular toxoplasmosis: A systematic review', *Infect Genet Evol*, 44: 199-209.

Ferreira Ade, M., R. W. Vitor, A. C. Carneiro, G. P. Brandao, and M. N. Melo. 2004. 'Genetic variability of Brazilian Toxoplasma gondii strains detected by random amplified polymorphic DNA-polymerase chain reaction (RAPD-PCR) and simple sequence repeat anchored-PCR (SSR-PCR)', *Infect Genet Evol*, 4: 131-42.

Ferreira, I. M., J. E. Vidal, T. A. Costa-Silva, C. S. Meira, R. M. Hiramoto, A. C. Penalva de Oliveira, and V. L. Pereira-Chioccola. 2008. 'Toxoplasma gondii: genotyping of strains from Brazilian AIDS patients with cerebral toxoplasmosis by multilocus PCR-RFLP markers', *Exp Parasitol*, 118: 221-7.

Figueiró-Filho, Ernesto Antonio, Alessandro Henrique Antunes Lopes, Flávio Renato de Almeida Senefonte, VG de Souza Júnior, Carlos Augusto Botelho, Mauro Silvério Figueiredo, and Geraldo Duarte. 2005. 'Toxoplasmose aguda: estudo da freqüência, taxa

de transmissão vertical e relação entre os testes diagnósticos materno-fetais em gestantes em estado da Região Centro-Oeste do Brasil', *Rev Bras Ginecol Obstet*, 27: 442-9.
Fleckenstein, M. C., M. L. Reese, S. Konen-Waisman, J. C. Boothroyd, J. C. Howard, and T. Steinfeldt. 2012. 'A Toxoplasma gondii pseudokinase inhibits host IRG resistance proteins', *PLoS Biol*, 10: e1001358.
Fletcher, S. 1965. 'Indirect Fluorescent Antibody Technique in the Serology of Toxoplasma Gondii', *J Clin Pathol*, 18: 193-9.
Fox, B. A., L. M. Rommereim, R. B. Guevara, A. Falla, M. A. Hortua Triana, Y. Sun, and D. J. Bzik. 2016. 'The Toxoplasma gondii Rhoptry Kinome Is Essential for Chronic Infection', *MBio*, 7.
Freppel, W., P. H. Puech, D. J. Ferguson, N. Azas, J. P. Dubey, and A. Dumetre. 2016. 'Macrophages facilitate the excystation and differentiation of Toxoplasma gondii sporozoites into tachyzoites following oocyst internalisation', *Sci Rep*, 6: 33654.
Freyre, A., and J. Falcon. 2004. 'Massive excystation of Toxoplasma gondii sporozoites', *Exp Parasitol*, 107: 72-7.
Fritz, H. M., K. R. Buchholz, X. Chen, B. Durbin-Johnson, D. M. Rocke, P. A. Conrad, and J. C. Boothroyd. 2012. 'Transcriptomic analysis of toxoplasma development reveals many novel functions and structures specific to sporozoites and oocysts', *PLoS One*, 7: e29998.
Fux, B., J. Nawas, A. Khan, D. B. Gill, C. Su, and L. D. Sibley. 2007. 'Toxoplasma gondii strains defective in oral transmission are also defective in developmental stage differentiation', *Infection and Immunity*, 75: 2580-90.
Gajria, B., A. Bahl, J. Brestelli, J. Dommer, S. Fischer, X. Gao, M. Heiges, J. Iodice, J. C. Kissinger, A. J. Mackey, D. F. Pinney, D. S. Roos, C. J. Stoeckert, Jr., H. Wang, and B. P. Brunk. 2008. 'ToxoDB: an integrated Toxoplasma gondii database resource', *Nucleic Acids Res*, 36: D553-6.
Gay, G., L. Braun, M. P. Brenier-Pinchart, J. Vollaire, V. Josserand, R. L. Bertini, A. Varesano, B. Touquet, P. J. De Bock, Y. Coute, I. Tardieux, A. Bougdour, and M. A. Hakimi. 2016. 'Toxoplasma gondii TgIST co-opts host chromatin repressors dampening STAT1-dependent gene regulation and IFN-gamma-mediated host defenses', *Journal of Experimental Medicine*, 213: 1779-98.
Gazzinelli, R. T., R. Mendonca-Neto, J. Lilue, J. Howard, and A. Sher. 2014. 'Innate Resistance against Toxoplasma gondii: An Evolutionary Tale of Mice, Cats, and Men', *Cell Host Microbe*, 15: 132-38.
Gibson, A. K., S. Raverty, D. M. Lambourn, J. Huggins, S. L. Magargal, and M. E. Grigg. 2011. 'Polyparasitism is associated with increased disease severity in Toxoplasma gondii-infected marine sentinel species', *PLoS Negl Trop Dis*, 5: e1142.
Gorfu, G., K. M. Cirelli, M. B. Melo, K. Mayer-Barber, D. Crown, B. H. Koller, S. Masters, A. Sher, S. H. Leppla, M. Moayeri, J. P. Saeij, and M. E. Grigg. 2014. 'Dual role for inflammasome sensors NLRP1 and NLRP3 in murine resistance to Toxoplasma gondii', *MBio*, 5.
Grigg, M. E., S. Bonnefoy, A. B. Hehl, Y. Suzuki, and J. C. Boothroyd. 2001. 'Success and virulence in toxoplasma as the result of sexual recombination between two distinct ancestries', *Science*, 294: 161-65.
Grigg, M. E., J. P. Dubey, and R. B. Nussenblatt. 2015. 'Ocular toxoplasmosis: lessons from Brazil', *Am J Ophthalmol*, 159: 999-1001.

Grigg, M. E., J. Ganatra, J. C. Boothroyd, and T. P. Margolis. 2001. 'Unusual abundance of atypical strains associated with human ocular toxoplasmosis', *J Infect Dis*, 184: 633-9.
Grigg, M. E., and N. Sundar. 2009. 'Sexual recombination punctuated by outbreaks and clonal expansions predicts Toxoplasma gondii population genetics', *Int J Parasitol*, 39: 925-33.
Gross, U., M. C. Kempf, F. Seeber, C. G. Luder, R. Lugert, and W. Bohne. 1997. 'Reactivation of chronic toxoplasmosis: is there a link to strain-specific differences in the parasite?', *Behring Inst Mitt*: 97-106.
Guo, M., J. P. Dubey, D. Hill, R. L. Buchanan, H. R. Gamble, J. L. Jones, and A. K. Pradhan. 2015. 'Prevalence and risk factors for Toxoplasma gondii infection in meat animals and meat products destined for human consumption', *J Food Prot*, 78: 457-76.
Guo, M., A. Mishra, R. L. Buchanan, J. P. Dubey, D. E. Hill, H. R. Gamble, J. L. Jones, and A. K. Pradhan. 2016. 'A Systematic Meta-Analysis of Toxoplasma gondii Prevalence in Food Animals in the United States', *Foodborne Pathog Dis*, 13: 109-18.
Hakansson, S., A. J. Charron, and L. D. Sibley. 2001. 'Toxoplasma evacuoles: a two-step process of secretion and fusion forms the parasitophorous vacuole', *EMBO J*, 20: 3132-44.
Hakimi, M. A., P. Olias, and L. D. Sibley. 2017. 'Toxoplasma Effectors Targeting Host Signaling and Transcription', *Clin Microbiol Rev*, 30: 615-45.
Harrell, M., and P. E. Carvounis. 2014. 'Current treatment of toxoplasma retinochoroiditis: an evidence-based review', *J Ophthalmol*, 2014: 273506.
Hehl, A. B., W. U. Basso, C. Lippuner, C. Ramakrishnan, M. Okoniewski, R. A. Walker, M. E. Grigg, N. C. Smith, and P. Deplazes. 2015. 'Asexual expansion of Toxoplasma gondii merozoites is distinct from tachyzoites and entails expression of non-overlapping gene families to attach, invade, and replicate within feline enterocytes', *Bmc Genomics*, 16: 66.
Heitman, J. 2010. 'Evolution of eukaryotic microbial pathogens via covert sexual reproduction', *Cell Host Microbe*, 8: 86-99.
Hermanns, T., U. B. Muller, S. Konen-Waisman, J. C. Howard, and T. Steinfeldt. 2015. 'The Toxoplasma gondii rhoptry protein ROP18 is an Irga6-specific kinase and regulated by the dense granule protein GRA7', *Cell Microbiol*.
Herrmann, D. C., P. Maksimov, A. Hotop, U. Gross, W. Daubener, O. Liesenfeld, U. Pleyer, F. J. Conraths, and G. Schares. 2014. 'Genotyping of samples from German patients with ocular, cerebral and systemic toxoplasmosis reveals a predominance of Toxoplasma gondii type II', *Int J Med Microbiol*, 304: 911-6.
Hide, G. 2016. 'Role of vertical transmission of Toxoplasma gondii in prevalence of infection', *Expert Rev Anti Infect Ther*, 14: 335-44.
Honnold, S. P., R. Braun, D. P. Scott, C. Sreekumar, and J. P. Dubey. 2005. 'Toxoplasmosis in a Hawaiian monk seal (Monachus schauinslandi)', *J Parasitol*, 91: 695-7.
Howe, D. K., and L. D. Sibley. 1994. 'Toxoplasma gondii: analysis of different laboratory stocks of the RH strain reveals genetic heterogeneity', *Exp Parasitol*, 78: 242-5.
———. 1995. 'Toxoplasma gondii comprises three clonal lineages: correlation of parasite genotype with human disease', *J Infect Dis*, 172: 1561-6.
Hunn, J. P., C. G. Feng, A. Sher, and J. C. Howard. 2011. 'The immunity-related GTPases in mammals: a fast-evolving cell-autonomous resistance system against intracellular pathogens', *Mamm Genome*, 22: 43-54.
Hunter, C. A., and L. D. Sibley. 2012. 'Modulation of innate immunity by Toxoplasma gondii virulence effectors', *Nat Rev Microbiol*, 10: 766-78.

Huson, D. H., and D. Bryant. 2006. 'Application of phylogenetic networks in evolutionary studies', *Mol Biol Evol*, 23: 254-67.

Iantorno, S. A., C. Durrant, A. Khan, M. J. Sanders, S. M. Beverley, W. C. Warren, M. Berriman, D. L. Sacks, J. A. Cotton, and M. E. Grigg. 2017. 'Gene Expression in Leishmania Is Regulated Predominantly by Gene Dosage', *MBio*, 8.

Inbar, E., N. S. Akopyants, M. Charmoy, A. Romano, P. Lawyer, D. E. Elnaiem, F. Kauffmann, M. Barhoumi, M. Grigg, K. Owens, M. Fay, D. E. Dobson, J. Shaik, S. M. Beverley, and D. Sacks. 2013. 'The mating competence of geographically diverse Leishmania major strains in their natural and unnatural sand fly vectors', *PLoS Genet*, 9: e1003672.

Jensen, K. D., A. Camejo, M. B. Melo, C. Cordeiro, L. Julien, G. M. Grotenbreg, E. M. Frickel, H. L. Ploegh, L. Young, and J. P. Saeij. 2015. 'Toxoplasma gondii Superinfection and Virulence during Secondary Infection Correlate with the Exact ROP5/ROP18 Allelic Combination', *MBio*, 6.

Jensen, K. D., K. Hu, R. J. Whitmarsh, M. A. Hassan, L. Julien, D. Lu, L. Chen, C. A. Hunter, and J. P. Saeij. 2013. 'Toxoplasma gondii Rhoptry 16 Kinase Promotes Host Resistance to Oral Infection and Intestinal Inflammation Only in the Context of the Dense Granule Protein GRA15', *Infection and Immunity*, 81: 2156-67.

Jensen, K. D., Y. Wang, E. D. Wojno, A. J. Shastri, K. Hu, L. Cornel, E. Boedec, Y. C. Ong, Y. H. Chien, C. A. Hunter, J. C. Boothroyd, and J. P. Saeij. 2011. 'Toxoplasma polymorphic effectors determine macrophage polarization and intestinal inflammation', *Cell Host Microbe*, 9: 472-83.

Jones, J. L., and J. P. Dubey. 2010. 'Waterborne toxoplasmosis--recent developments', *Exp Parasitol*, 124: 10-25.

Joy, D. A., X. Feng, J. Mu, T. Furuya, K. Chotivanich, A. U. Krettli, M. Ho, A. Wang, N. J. White, E. Suh, P. Beerli, and X. Z. Su. 2003. 'Early origin and recent expansion of Plasmodium falciparum', *Science*, 300: 318-21.

Khairat, R., M. Ball, C. C. Chang, R. Bianucci, A. G. Nerlich, M. Trautmann, S. Ismail, G. M. Shanab, A. M. Karim, Y. Z. Gad, and C. M. Pusch. 2013. 'First insights into the metagenome of Egyptian mummies using next-generation sequencing', *J Appl Genet*, 54: 309-25.

Khan, A., M. S. Behnke, I. R. Dunay, M. W. White, and L. D. Sibley. 2009. 'Phenotypic and gene expression changes among clonal type I strains of Toxoplasma gondii', *Eukaryot Cell*, 8: 1828-36.

Khan, A., U. Bohme, K. A. Kelly, E. Adlem, K. Brooks, M. Simmonds, K. Mungall, M. A. Quail, C. Arrowsmith, T. Chillingworth, C. Churcher, D. Harris, M. Collins, N. Fosker, A. Fraser, Z. Hance, K. Jagels, S. Moule, L. Murphy, S. O'Neil, M. A. Rajandream, D. Saunders, K. Seeger, S. Whitehead, T. Mayr, X. Xuan, J. Watanabe, Y. Suzuki, H. Wakaguri, S. Sugano, C. Sugimoto, I. Paulsen, A. J. Mackey, D. S. Roos, N. Hall, M. Berriman, B. Barrell, L. D. Sibley, and J. W. Ajioka. 2006. 'Common inheritance of chromosome Ia associated with clonal expansion of Toxoplasma gondii', *Genome Res*, 16: 1119-25.

Khan, A., J. P. Dubey, C. Su, J. W. Ajioka, B. M. Rosenthal, and L. D. Sibley. 2011. 'Genetic analyses of atypical Toxoplasma gondii strains reveal a fourth clonal lineage in North America', *Int J Parasitol*, 41: 645-55.

Khan, A., N. Miller, D. S. Roos, J. P. Dubey, D. Ajzenberg, M. L. Darde, J. W. Ajioka, B. Rosenthal, and L. D. Sibley. 2011. 'A monomorphic haplotype of chromosome Ia is

associated with widespread success in clonal and nonclonal populations of Toxoplasma gondii', *MBio*, 2: e00228-11.
Khan, A., J. S. Shaik, M. Behnke, Q. L. Wang, J. P. Dubey, H. A. Lorenzi, J. W. Ajioka, B. M. Rosenthal, and L. D. Sibley. 2014. 'NextGen sequencing reveals short double crossovers contribute disproportionately to genetic diversity in Toxoplasma gondii', *Bmc Genomics*, 15.
Khan, A., S. Taylor, J. W. Ajioka, B. M. Rosenthal, and L. D. Sibley. 2009. 'Selection at a single locus leads to widespread expansion of Toxoplasma gondii lineages that are virulent in mice', *PLoS Genet*, 5: e1000404.
Khan, A., S. Taylor, C. Su, A. J. Mackey, J. Boyle, R. Cole, D. Glover, K. Tang, I. T. Paulsen, M. Berriman, J. C. Boothroyd, E. R. Pfefferkorn, J. P. Dubey, J. W. Ajioka, D. S. Roos, J. C. Wootton, and L. D. Sibley. 2005. 'Composite genome map and recombination parameters derived from three archetypal lineages of Toxoplasma gondii', *Nucleic Acids Res*, 33: 2980-92.
Kim, K. 2006. 'Time to screen for congenital toxoplasmosis?', *Clin Infect Dis*, 42: 1395-7.
Koblansky, A. A., D. Jankovic, H. Oh, S. Hieny, W. Sungnak, R. Mathur, M. S. Hayden, S. Akira, A. Sher, and S. Ghosh. 2013. 'Recognition of profilin by Toll-like receptor 12 is critical for host resistance to Toxoplasma gondii', *Immunity*, 38: 119-30.
Kong, J. T., M. E. Grigg, L. Uyetake, S. Parmley, and J. C. Boothroyd. 2003. 'Serotyping of Toxoplasma gondii infections in humans using synthetic peptides', *Journal of Infectious Diseases*, 187: 1484-95.
Kreuder, C., M. A. Miller, D. A. Jessup, L. J. Lowenstine, M. D. Harris, J. A. Ames, T. E. Carpenter, P. A. Conrad, and J. A. Mazet. 2003. 'Patterns of mortality in southern sea otters (Enhydra lutris nereis) from 1998-2001', *J Wildl Dis*, 39: 495-509.
Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. 'Clustal W and Clustal X version 2.0', *Bioinformatics*, 23: 2947-8.
Lehmann, T., P. L. Marcet, D. H. Graham, E. R. Dahl, and J. P. Dubey. 2006. 'Globalization and the population structure of Toxoplasma gondii', *Proc Natl Acad Sci U S A*, 103: 11423-8.
Li, X. L., H. X. Wei, H. Zhang, H. J. Peng, and D. S. Lindsay. 2014. 'A meta analysis on risks of adverse pregnancy outcomes in Toxoplasma gondii infection', *PLoS One*, 9: e97775.
Lilue, J., U. B. Muller, T. Steinfeldt, and J. C. Howard. 2013. 'Reciprocal virulence and resistance polymorphism in the relationship between Toxoplasma gondii and the house mouse', *Elife*, 2: e01298.
Lindsay, D. S., B. L. Blagburn, and J. P. Dubey. 2002. 'Survival of nonsporulated Toxoplasma gondii oocysts under refrigerator conditions', *Vet Parasitol*, 103: 309-13.
Lindsay, D. S., and J. P. Dubey. 2009. 'Long-term survival of Toxoplasma gondii sporulated oocysts in seawater', *J Parasitol*, 95: 1019-20.
Liu, Q., L. Singla, and H. Zhou. 2012. 'Vaccines against Toxoplasma gondii: Status, challenges and future directions', *Hum Vaccin Immunother*, 8: 1305-8.
Lopes, A. P., J. P. Dubey, M. L. Darde, and L. Cardoso. 2014. 'Epidemiological review of Toxoplasma gondii infection in humans and animals in Portugal', *Parasitology*, 141: 1699-708.
Lorenzi, H., A. Khan, M. S. Behnke, S. Namasivayam, L. S. Swapna, M. Hadjithomas, S. Karamycheva, D. Pinney, B. P. Brunk, J. W. Ajioka, D. Ajzenberg, J. C. Boothroyd, J. P. Boyle, M. L. Darde, M. A. Diaz-Miranda, J. P. Dubey, H. M. Fritz, S. M. Gennari, B. D.

Gregory, K. Kim, J. P. Saeij, C. Su, M. W. White, X. Q. Zhu, D. K. Howe, B. M. Rosenthal, M. E. Grigg, J. Parkinson, L. Liu, J. C. Kissinger, D. S. Roos, and L. David Sibley. 2016. 'Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic Toxoplasma gondii genomes', *Nat Commun*, 7: 10147.

Lykins, J., K. Wang, K. Wheeler, F. Clouser, A. Dixon, K. El Bissati, Y. Zhou, C. Lyttle, A. Rzhetsky, and R. McLeod. 2016. 'Understanding Toxoplasmosis in the United States Through "Large Data" Analyses', *Clin Infect Dis*, 63: 468-75.

Ma, J. S., M. Sasai, J. Ohshima, Y. Lee, H. Bando, K. Takeda, and M. Yamamoto. 2014. 'Selective and strain-specific NFAT4 activation by the Toxoplasma gondii polymorphic dense granule protein GRA6', *Journal of Experimental Medicine*, 211: 2013-32.

Mackinnon, M. J., and A. F. Read. 2004. 'Virulence in malaria: an evolutionary viewpoint', *Philos Trans R Soc Lond B Biol Sci*, 359: 965-86.

Matrajt, M., S. O. Angel, V. Pszenny, E. Guarnera, D. S. Roos, and J. C. Garberi. 1999. 'Arrays of repetitive DNA elements in the largest chromosomes of Toxoplasma gondii', *Genome*, 42: 265-9.

Matsuzaki, M., T. Kikuchi, K. Kita, S. Kojima, and T. Kuroiwa. 2001. 'Large amounts of apicoplast nucleoid DNA and its segregation in Toxoplasma gondii', *Protoplasma*, 218: 180-91.

McLeod, R., K. M. Boyer, D. Lee, E. Mui, K. Wroblewski, T. Karrison, A. G. Noble, S. Withers, C. N. Swisher, P. T. Heydemann, M. Sautter, J. Babiarz, P. Rabiah, P. Meier, M. E. Grigg, and Group Toxoplasmosis Study. 2012. 'Prematurity and severity are associated with Toxoplasma gondii alleles (NCCCTS, 1981-2009)', *Clin Infect Dis*, 54: 1595-605.

Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe. 1999. 'Food-related illness and death in the United States', *Emerg Infect Dis*, 5: 607-25.

Meissner, M., M. S. Breinich, P. R. Gilson, and B. S. Crabb. 2007. 'Molecular genetic tools in Toxoplasma and Plasmodium: achievements and future needs', *Curr Opin Microbiol*, 10: 349-56.

Melo, M. B., Q. P. Nguyen, C. Cordeiro, M. A. Hassan, N. Yang, R. McKell, E. E. Rosowski, L. Julien, V. Butty, M. L. Darde, D. Ajzenberg, K. Fitzgerald, L. H. Young, and J. P. Saeij. 2013. 'Transcriptional analysis of murine macrophages infected with different Toxoplasma strains identifies novel regulation of host signaling pathways', *PLoS Pathog*, 9: e1003779.

Miller, M. A., I. A. Gardner, A. Packham, J. K. Mazet, K. D. Hanni, D. Jessup, J. Estes, R. Jameson, E. Dodd, B. C. Barr, L. J. Lowenstine, F. M. Gulland, and P. A. Conrad. 2002. 'Evaluation of an Indirect Fluorescent Antibody Test (Ifat) for Demonstration of Antibodies to Toxoplasma Gondii in the Sea Otter (Enhydra Lutris)', *Journal of Parasitology*, 88: 594-99.

Miller, M. A., M. E. Grigg, C. Kreuder, E. R. James, A. C. Melli, P. R. Crosbie, D. A. Jessup, J. C. Boothroyd, D. Brownstein, and P. A. Conrad. 2004. 'An unusual genotype of Toxoplasma gondii is common in California sea otters (Enhydra lutris nereis) and is a cause of mortality', *Int J Parasitol*, 34: 275-84.

Miller, M. A., W. A. Miller, P. A. Conrad, E. R. James, A. C. Melli, C. M. Leutenegger, H. A. Dabritz, A. E. Packham, D. Paradies, M. Harris, J. Ames, D. A. Jessup, K. Worcester, and M. E. Grigg. 2008. 'Type X Toxoplasma gondii in a wild mussel and terrestrial

carnivores from coastal California: new linkages between terrestrial mammals, runoff and toxoplasmosis of sea otters', *Int J Parasitol*, 38: 1319-28.
Minot, S., M. B. Melo, F. Li, D. Lu, W. Niedelman, S. S. Levine, and J. P. J. Saeij. 2012. 'Admixture and recombination among Toxoplasma gondii lineages explain global genome diversity', *Proc Natl Acad Sci U S A*, 109: 13458-63.
Miotto, O., J. Almagro-Garcia, M. Manske, B. Macinnis, S. Campino, K. A. Rockett, C. Amaratunga, P. Lim, S. Suon, S. Sreng, J. M. Anderson, S. Duong, C. Nguon, C. M. Chuor, D. Saunders, Y. Se, C. Lon, M. M. Fukuda, L. Amenga-Etego, A. V. Hodgson, V. Asoala, M. Imwong, S. Takala-Harrison, F. Nosten, X. Z. Su, P. Ringwald, F. Ariey, C. Dolecek, T. T. Hien, M. F. Boni, C. Q. Thai, A. Amambua-Ngwa, D. J. Conway, A. A. Djimde, O. K. Doumbo, I. Zongo, J. B. Ouedraogo, D. Alcock, E. Drury, S. Auburn, O. Koch, M. Sanders, C. Hubbart, G. Maslen, V. Ruano-Rubio, D. Jyothi, A. Miles, J. O'Brien, C. Gamble, S. O. Oyola, J. C. Rayner, C. I. Newbold, M. Berriman, C. C. Spencer, G. McVean, N. P. Day, N. J. White, D. Bethell, A. M. Dondorp, C. V. Plowe, R. M. Fairhurst, and D. P. Kwiatkowski. 2013. 'Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia', *Nat Genet*, 45: 648-55.
Molina-Cruz, Alvaro, and Carolina Barillas-Mury. 2014. 'The remarkable journey of adaptation of the Plasmodium falciparum malaria parasite to New World anopheline mosquitoes', *Mem Inst Oswaldo Cruz*, 109: 662-67.
Mordue, D. G., N. Desai, M. Dustin, and L. D. Sibley. 1999. 'Invasion by Toxoplasma gondii establishes a moving junction that selectively excludes host cell plasma membrane proteins on the basis of their membrane anchoring', *Journal of Experimental Medicine*, 190: 1783-92.
Ni, M., M. Feretzaki, S. Sun, X. Wang, and J. Heitman. 2011. 'Sex in fungi', *Annu Rev Genet*, 45: 405-30.
Niedelman, W., D. A. Gold, E. E. Rosowski, J. K. Sprokholt, D. Lim, A. Farid Arenas, M. B. Melo, E. Spooner, M. B. Yaffe, and J. P. Saeij. 2012. 'The rhoptry proteins ROP18 and ROP5 mediate Toxoplasma gondii evasion of the murine, but not the human, interferon-gamma response', *PLoS Pathog*, 8: e1002784.
Olias, P., R. D. Etheridge, Y. Zhang, M. J. Holtzman, and L. D. Sibley. 2016. 'Toxoplasma Effector Recruits the Mi-2/NuRD Complex to Repress STAT1 Transcription and Block IFN-gamma-Dependent Gene Expression', *Cell Host Microbe*, 20: 72-82.
Pan, S., R. C. Thompson, M. E. Grigg, N. Sundar, A. Smith, and A. J. Lymbery. 2012. 'Western Australian marsupials are multiply infected with genetically diverse strains of Toxoplasma gondii', *PLoS One*, 7: e45147.
Parameswaran, N., R. C. Thompson, N. Sundar, S. Pan, M. Johnson, N. C. Smith, and M. E. Grigg. 2010. 'Non-archetypal Type II-like and atypical strains of Toxoplasma gondii infecting marsupials of Australia', *Int J Parasitol*, 40: 635-40.
Paya, E., I. Noemi, R. Tassara, P. Catalan, and C. L. Aviles. 2012. '[Prophylaxis against Toxoplasma gondii disease in pediatric and adult patients undergoing solid organ and hematopoietic stem cells transplantation]', *Rev Chilena Infectol*, 29 Suppl 1: S37-9.
Pearson, T. A., and T. A. Manolio. 2008. 'How to interpret a genome-wide association study', *Jama-Journal of the American Medical Association*, 299: 1335-44.
Peixoto, L., F. Chen, O. S. Harb, P. H. Davis, D. P. Beiting, C. S. Brownback, D. Ouloguem, and D. S. Roos. 2010. 'Integrative genomic approaches highlight a family of parasite-specific kinases that regulate host responses', *Cell Host Microbe*, 8: 208-18.

Pellegrini, E., A. Palencia, L. Braun, U. Kapp, A. Bougdour, H. Belrhali, M. W. Bowler, and M. A. Hakimi. 2017. 'Structural Basis for the Subversion of MAP Kinase Signaling by an Intrinsically Disordered Parasite Secreted Agonist', *Structure*, 25: 16-26.

Penman, B., C. Buckee, S. Gupta, and S. Nee. 2010. 'Genome-wide association studies in Plasmodium species', *Bmc Biology*, 8.

Pernas, L., Y. Adomako-Ankomah, A. J. Shastri, S. E. Ewald, M. Treeck, J. P. Boyle, and J. C. Boothroyd. 2014. 'Toxoplasma effector MAF1 mediates recruitment of host mitochondria and impacts the host response', *PLoS Biol*, 12: e1001845.

Pfefferkorn, E. R. 1981. 'Toxoplasmagondii and the biochemistry of intracellular parasitism', *TIBS*.

Pfefferkorn, E. R., and L. C. Pfefferkorn. 1979. 'Quantitative studies of the mutagenesis of Toxoplasma gondii', *J Parasitol*, 65: 364-70.

Pfefferkorn, E. R., L. C. Pfefferkorn, and E. D. Colby. 1977. 'Development of gametes and oocysts in cats fed cysts derived from cloned trophozoites of Toxoplasma gondii', *J Parasitol*, 63: 158-9.

Pfefferkorn, L. C., and E. R. Pfefferkorn. 1980. 'Toxoplasma gondii: genetic recombination between drug resistant mutants', *Exp Parasitol*, 50: 305-16.

Piergili Fioretti, D. 2004. '[Problems and limitations of conventional and innovative methods for the diagnosis of Toxoplasmosis in humans and animals]', *Parassitologia*, 46: 177-81.

Pszenny, V., S. O. Angel, V. G. Duschak, M. Paulino, B. Ledesma, M. I. Yabo, E. Guarnera, A. M. Ruiz, and E. J. Bontempi. 2000. 'Molecular cloning, sequencing and expression of a serine proteinase inhibitor gene from Toxoplasma gondii', *Mol Biochem Parasitol*, 107: 241-9.

Quinlan, A. R., and I. M. Hall. 2010. 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26: 841-2.

Rajendran, C., C. Su, and J. P. Dubey. 2012. 'Molecular genotyping of Toxoplasma gondii from Central and South America revealed high diversity within and between populations', *Infect Genet Evol*, 12: 359-68.

Ramirez, J. D., and M. S. Llewellyn. 2014. 'Reproductive clonality in protozoan pathogens-truth or artefact?', *Mol Ecol*, 23: 4195-202.

Read, A. F., M. Anwar, D. Shutler, and S. Nee. 1995. 'Sex allocation and population structure in malaria and related parasitic protozoa', *Proc Biol Sci*, 260: 359-63.

Rich, S. M., M. C. Licht, R. R. Hudson, and F. J. Ayala. 1998. 'Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of Plasmodium falciparum', *Proc Natl Acad Sci U S A*, 95: 4425-30.

Rico-Torres, C. P., J. A. Vargas-Villavicencio, and D. Correa. 2016. 'Is Toxoplasma gondii type related to clinical outcome in human congenital infection? Systematic and critical review', *Eur J Clin Microbiol Infect Dis*, 35: 1079-88.

Roach, K. C., and J. Heitman. 2014. 'Unisexual reproduction reverses Muller's ratchet', *Genetics*, 198: 1059-69.

Rogers, M. B., T. Downing, B. A. Smith, H. Imamura, M. Sanders, M. Svobodova, P. Volf, M. Berriman, J. A. Cotton, and D. F. Smith. 2014. 'Genomic confirmation of hybridisation and recent inbreeding in a vector-isolated leishmania population', *PLoS Genet*, 10: e1004092.

Roiko, M. S., N. Svezhova, and V. B. Carruthers. 2014. 'Acidification Activates Toxoplasma gondii Motility and Egress by Enhancing Protein Secretion and Cytolytic Activity', *PLoS Pathog*, 10: e1004488.

Romano, A., E. Inbar, A. Debrabant, M. Charmoy, P. Lawyer, F. Ribeiro-Gomes, M. Barhoumi, M. Grigg, J. Shaik, D. Dobson, S. M. Beverley, and D. L. Sacks. 2014. 'Cross-species genetic exchange between visceral and cutaneous strains of Leishmania in the sand fly vector', *Proc Natl Acad Sci U S A*, 111: 16808-13.

Rosowski, E. E., D. Lu, L. Julien, L. Rodda, R. A. Gaiser, K. D. Jensen, and J. P. Saeij. 2011. 'Strain-specific activation of the NF-kappaB pathway by GRA15, a novel Toxoplasma gondii dense granule protein', *Journal of Experimental Medicine*, 208: 195-212.

Ryu, J. S., D. Y. Min, M. H. Ahn, H. G. Choi, S. C. Rho, Y. J. Shin, B. Choi, and H. D. Joo. 1996. 'Toxoplasma antibody titers by ELISA and indirect latex agglutination test in pregnant women', *Korean J Parasitol*, 34: 233-8.

Saeij, J. P., J. P. Boyle, and J. C. Boothroyd. 2005. 'Differences among the three major strains of Toxoplasma gondii and their specific interactions with the infected host', *Trends Parasitol*, 21: 476-81.

Saeij, J. P., J. P. Boyle, S. Coller, S. Taylor, L. D. Sibley, E. T. Brooke-Powell, J. W. Ajioka, and J. C. Boothroyd. 2006. 'Polymorphic secreted kinases are key virulence factors in toxoplasmosis', *Science*, 314: 1780-3.

Saeij, J. P., J. P. Boyle, M. E. Grigg, G. Arrizabalaga, and J. C. Boothroyd. 2005. 'Bioluminescence imaging of Toxoplasma gondii infection in living mice reveals dramatic differences between strains', *Infection and Immunity*, 73: 695-702.

Saeij, J. P., S. Coller, J. P. Boyle, M. E. Jerome, M. W. White, and J. C. Boothroyd. 2007. 'Toxoplasma co-opts host gene expression by injection of a polymorphic kinase homologue', *Nature*, 445: 324-7.

Schluter, D., W. Daubener, G. Schares, U. Gross, U. Pleyer, and C. Luder. 2014. 'Animals are key to human toxoplasmosis', *Int J Med Microbiol*, 304: 917-29.

Schneider, G., U. Dobrindt, B. Middendorf, B. Hochhut, V. Szijarto, L. Emody, and J. Hacker. 2011. 'Mobilisation and remobilisation of a large archetypal pathogenicity island of uropathogenic Escherichia coli in vitro support the role of conjugation for horizontal transfer of genomic islands', *BMC Microbiol*, 11: 210.

Schoondermarkvandeven, E., J. Galama, C. Kraaijeveld, J. Vandruten, J. Meuwissen, and W. Melchers. 1993. 'Value of the Polymerase Chain-Reaction for the Detection of Toxoplasma-Gondii in Cerebrospinal-Fluid from Patients with Aids', *Clinical Infectious Diseases*, 16: 661-66.

Sedwick, C. 2014. 'MAF1 sequesters mitochondria for Toxoplasma gondii', *PLoS Biol*, 12: e1001846.

Seeber, F., and S. Steinfelder. 2016. 'Recent advances in understanding apicomplexan parasites', *F1000Res*, 5.

Selleck, E. M., S. J. Fentress, W. L. Beatty, D. Degrandi, K. Pfeffer, H. W. th Virgin, J. D. Macmicking, and L. D. Sibley. 2013. 'Guanylate-binding protein 1 (Gbp1) contributes to cell-autonomous immunity against Toxoplasma gondii', *PLoS Pathog*, 9: e1003320.

Selmecki, A. M., Y. E. Maruvka, P. A. Richmond, M. Guillet, N. Shoresh, A. L. Sorenson, S. De, R. Kishony, F. Michor, R. Dowell, and D. Pellman. 2015. 'Polyploidy can drive rapid adaptation in yeast', *Nature*, 519: 349-52.

Shaik, J. S., A. Khan, S. M. Beverley, and L. D. Sibley. 2015. 'REDHORSE-REcombination and Double crossover detection in Haploid Organisms using next-geneRation SEquencing data', *Bmc Genomics*, 16: 133.

Sharma, P., and C. E. Chitnis. 2013. 'Key molecular events during host cell invasion by Apicomplexan pathogens', *Curr Opin Microbiol*.

Shastri, A. J., N. D. Marino, M. Franco, M. B. Lodoen, and J. C. Boothroyd. 2014. 'GRA25 Is a Novel Virulence Factor of Toxoplasma gondii and Influences the Host Immune Response', *Infection and Immunity*, 82: 2595-605.

Shobab, L., U. Pleyer, J. Johnsen, S. Metzner, E. R. James, N. Torun, M. P. Fay, O. Liesenfeld, and M. E. Grigg. 2013. 'Toxoplasma Serotype Is Associated With Development of Ocular Toxoplasmosis.' in, *J Infect Dis*.

Shwab, E. K., T. Jiang, H. F. Pena, S. M. Gennari, J. P. Dubey, and C. Su. 2016. 'The ROP18 and ROP5 gene allele types are highly predictive of virulence in mice across globally distributed strains of Toxoplasma gondii', *Int J Parasitol*, 46: 141-6.

Shwab, E. K., X. Q. Zhu, D. Majumdar, H. F. Pena, S. M. Gennari, J. P. Dubey, and C. Su. 2014. 'Geographical patterns of Toxoplasma gondii genetic diversity revealed by multilocus PCR-RFLP genotyping', *Parasitology*, 141: 453-61.

Sibley, L. D. 2009. 'Development of forward genetics in Toxoplasma gondii', *Int J Parasitol*, 39: 915-24.

Sibley, L. D., and J. W. Ajioka. 2008. 'Population structure of Toxoplasma gondii: clonal expansion driven by infrequent recombination and selective sweeps', *Annu Rev Microbiol*, 62: 329-51.

Sibley, L. D., and J. C. Boothroyd. 1992. 'Virulent strains of Toxoplasma gondii comprise a single clonal lineage', *Nature*, 359: 82-5.

Sibley, L. D., A. Khan, J. W. Ajioka, and B. M. Rosenthal. 2009. 'Genetic diversity of Toxoplasma gondii in animals and humans', *Philos Trans R Soc Lond B Biol Sci*, 364: 2749-61.

Sibley, L. D., A. J. LeBlanc, E. R. Pfefferkorn, and J. C. Boothroyd. 1992. 'Generation of a restriction fragment length polymorphism linkage map for Toxoplasma gondii', *Genetics*, 132: 1003-15.

Sidik, S. M., D. Huet, S. M. Ganesan, M. H. Huynh, T. Wang, A. S. Nasamu, P. Thiru, J. P. Saeij, V. B. Carruthers, J. C. Niles, and S. Lourido. 2016. 'A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes', *Cell*, 166: 1423-35 e12.

Silva, J. C. R., F. Ferreira, R. A. Dias, D. Ajzenberg, M. F. V. Marvulo, F. J. R. Magalhaes, Cdfl Filho, S. Oliveira, H. S. Soares, T. F. Feitosa, J. Aizawa, L. C. Alves, R. A. Mota, J. P. Dubey, S. M. Gennari, and H. F. J. Pena. 2017. 'Cat-rodent Toxoplasma gondii Type II-variant circulation and limited genetic diversity on the Island of Fernando de Noronha, Brazil', *Parasit Vectors*, 10: 220.

Silva, L. A., R. O. Andrade, A. C. Carneiro, and R. W. Vitor. 2014. 'Overlapping Toxoplasma gondii genotypes circulating in domestic animals and humans in Southeastern Brazil', *PLoS One*, 9: e90237.

Skariah, S., M. K. McIntyre, and D. G. Mordue. 2010. 'Toxoplasma gondii: determinants of tachyzoite to bradyzoite conversion', *Parasitol Res*, 107: 253-60.

Smith, R., K. Sheppard, K. DiPetrillo, and G. Churchill. 2009. 'Quantitative trait locus analysis using J/qtl', *Methods Mol Biol*, 573: 175-88.

Soete, M., B. Fortier, D. Camus, and J. F. Dubremetz. 1993. 'Toxoplasma gondii: kinetics of bradyzoite-tachyzoite interconversion in vitro', *Exp Parasitol*, 76: 259-64.
Soheilian, M., M. M. Sadoughi, M. Ghajarnia, M. H. Dehghan, S. Yazdani, H. Behboudi, A. Anisian, and G. A. Peyman. 2005. 'Prospective randomized trial of trimethoprim/sulfamethoxazole versus pyrimethamine and sulfadiazine in the treatment of ocular toxoplasmosis', *Ophthalmology*, 112: 1876-82.
Steinfeldt, T., S. Konen-Waisman, L. Tong, N. Pawlowski, T. Lamkemeyer, L. D. Sibley, J. P. Hunn, and J. C. Howard. 2010. 'Phosphorylation of mouse immunity-related GTPase (IRG) resistance proteins is an evasion strategy for virulent Toxoplasma gondii', *PLoS Biol*, 8: e1000576.
Striepen, B., M. J. Crawford, M. K. Shaw, L. G. Tilney, F. Seeber, and D. S. Roos. 2000. 'The plastid of Toxoplasma gondii is divided by association with the centrosomes', *Journal of Cell Biology*, 151: 1423-34.
Su, C., D. Evans, R. H. Cole, J. C. Kissinger, J. W. Ajioka, and L. D. Sibley. 2003. 'Recent expansion of Toxoplasma through enhanced oral transmission', *Science*, 299: 414-6.
Su, C., D. K. Howe, J. P. Dubey, J. W. Ajioka, and L. D. Sibley. 2002. 'Identification of quantitative trait loci controlling acute virulence in Toxoplasma gondii', *Proc Natl Acad Sci U S A*, 99: 10753-8.
Su, C., A. Khan, P. Zhou, D. Majumdar, D. Ajzenberg, M. L. Darde, X. Q. Zhu, J. W. Ajioka, B. M. Rosenthal, J. P. Dubey, and L. D. Sibley. 2012. 'Globally diverse Toxoplasma gondii isolates comprise six major clades originating from a small number of distinct ancestral lineages', *Proc Natl Acad Sci U S A*, 109: 5844-9.
Su, C., X. Zhang, and J. P. Dubey. 2006. 'Genotyping of Toxoplasma gondii by multilocus PCR-RFLP markers: a high resolution and simple method for identification of parasites', *Int J Parasitol*, 36: 841-8.
Sun, S., R. B. Billmyre, P. A. Mieczkowski, and J. Heitman. 2014. 'Unisexual reproduction drives meiotic recombination and phenotypic and karyotypic plasticity in Cryptococcus neoformans', *PLoS Genet*, 10: e1004849.
Sundar, N., R. A. Cole, N. J. Thomas, D. Majumdar, J. P. Dubey, and C. Su. 2008. 'Genetic diversity among sea otter isolates of Toxoplasma gondii', *Vet Parasitol*, 151: 125-32.
Suzuki, Y., M. A. Orellana, R. D. Schreiber, and J. S. Remington. 1988. 'Interferon-gamma: the major mediator of resistance against Toxoplasma gondii', *Science*, 240: 516-8.
Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. 'MEGA6: Molecular Evolutionary Genetics Analysis version 6.0', *Mol Biol Evol*, 30: 2725-9.
Taylor, S., A. Barragan, C. Su, B. Fux, S. J. Fentress, K. Tang, W. L. Beatty, H. E. Hajj, M. Jerome, M. S. Behnke, M. White, J. C. Wootton, and L. D. Sibley. 2006. 'A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen Toxoplasma gondii', *Science*, 314: 1776-80.
Tenter, A. M. 1995. 'Current research on Sarcocystis species of domestic animals', *Int J Parasitol*, 25: 1311-30.
Tenter, A. M., A. R. Heckeroth, and L. M. Weiss. 2000. 'Toxoplasma gondii: from animals to humans', *Int J Parasitol*, 30: 1217-58.
Thompson, R. C. 2013. 'Parasite zoonoses and wildlife: One Health, spillover and human activity', *Int J Parasitol*, 43: 1079-88.
Thompson, R. C., A. J. Lymbery, and A. Smith. 2010. 'Parasites, emerging disease and wildlife conservation', *Int J Parasitol*, 40: 1163-70.

Tibayrenc, M., and F. J. Ayala. 1991. 'Towards a population genetics of microorganisms: The clonal theory of parasitic protozoa', *Parasitol Today*, 7: 228-32.
———. 2002. 'The clonal theory of parasitic protozoa: 12 years on', *Trends Parasitol*, 18: 405-10.
———. 2014. 'New insights into clonality and panmixia in Plasmodium and toxoplasma', *Adv Parasitol*, 84: 253-68.
Tilley, M., M. E. Fichera, M. E. Jerome, D. S. Roos, and M. W. White. 1997. 'Toxoplasma gondii sporozoites form a transient parasitophorous vacuole that is impermeable and contains only a subset of dense-granule proteins', *Infection and Immunity*, 65: 4598-605.
Tonkin, M. L., M. Roques, M. H. Lamarque, M. Pugniere, D. Douguet, J. Crawford, M. Lebrun, and M. J. Boulanger. 2011. 'Host Cell Invasion by Apicomplexan Parasites: Insights from the Co-Structure of AMA1 with a RON2 Peptide', *Science*, 333: 463-67.
Torrey, E. F., and R. H. Yolken. 2013. 'Toxoplasma oocysts as a public health problem', *Trends Parasitol*, 29: 380-4.
Trevors, J.T. 1998. 'Review: Bacterial Population Genetics', *World Journal of Microbiology and Biotechnology*, 14: 1-5.
Trouplin, V., N. Boucherit, L. Gorvel, F. Conti, G. Mottola, and E. Ghigo. 2013. 'Bone marrow-derived macrophage production', *J Vis Exp*: e50966.
Twiddy, S. S., and E. C. Holmes. 2003. 'The extent of homologous recombination in members of the genus Flavivirus', *J Gen Virol*, 84: 429-40.
Vaishnava, S., and B. Striepen. 2006. 'The cell biology of secondary endosymbiosis--how parasites build, divide and segregate the apicoplast', *Mol Microbiol*, 61: 1380-7.
Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo. 2013. 'From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline', *Curr Protoc Bioinformatics*, 11: 11 10 1-11 10 33.
VanWormer, E., T. E. Carpenter, P. Singh, K. Shapiro, W. W. Wallender, P. A. Conrad, J. L. Largier, M. P. Maneta, and J. A. Mazet. 2016. 'Coastal development and precipitation drive pathogen flow from land to sea: evidence from a Toxoplasma gondii and felid host system', *Sci Rep*, 6: 29252.
VanWormer, E., M. A. Miller, P. A. Conrad, M. E. Grigg, D. Rejmanek, T. E. Carpenter, and J. A. Mazet. 2014. 'Using molecular epidemiology to track Toxoplasma gondii from terrestrial carnivores to marine hosts: implications for public health and conservation', *PLoS Negl Trop Dis*, 8: e2852.
Vaudaux, J. D., C. Muccioli, E. R. James, C. Silveira, S. L. Magargal, C. Jung, J. P. Dubey, J. L. Jones, M. Z. Doymaz, D. A. Bruckner, R. Belfort, Jr., G. N. Holland, and M. E. Grigg. 2010. 'Identification of an atypical strain of toxoplasma gondii as the cause of a waterborne outbreak of toxoplasmosis in Santa Isabel do Ivai, Brazil', *J Infect Dis*, 202: 1226-33.
Verma, Ramesh, and Pardeep Khanna. 2014. 'Development of Toxoplasma gondii vaccine', *Hum Vaccin Immunother*, 9: 291-93.
Verma, S. K., D. Ajzenberg, A. Rivera-Sanchez, C. Su, and J. P. Dubey. 2015. 'Genetic characterization of Toxoplasma gondii isolates from Portugal, Austria and Israel reveals higher genetic variability within the type II lineage', *Parasitology*, 142: 948-57.

Verma, S. K., A. R. Sweeny, M. J. Lovallo, R. Calero-Bernal, O. C. Kwok, T. Jiang, C. Su, M. E. Grigg, and J. P. Dubey. 2017. 'Seroprevalence, isolation and co-infection of multiple Toxoplasma gondii strains in individual bobcats (Lynx rufus) from Mississippi, USA', *Int J Parasitol*, 47: 297-303.

Vijaykrishna, D., G. J. Smith, O. G. Pybus, H. Zhu, S. Bhatt, L. L. Poon, S. Riley, J. Bahl, S. K. Ma, C. L. Cheung, R. A. Perera, H. Chen, K. F. Shortridge, R. J. Webby, R. G. Webster, Y. Guan, and J. S. Peiris. 2011. 'Long-term evolution and transmission dynamics of swine influenza A virus', *Nature*, 473: 519-22.

Vitaliano, S. N., H. S. Soares, A. H. Minervino, A. L. Santos, K. Werther, M. F. Marvulo, D. B. Siqueira, H. F. Pena, R. M. Soares, C. Su, and S. M. Gennari. 2014. 'Genetic characterization of Toxoplasma gondii from Brazilian wildlife revealed abundant new genotypes', *Int J Parasitol Parasites Wildl*, 3: 276-83.

Wakeman, K. C., M. B. Heintzelman, and B. S. Leander. 2014. 'Comparative ultrastructure and molecular phylogeny of Selenidium melongena n. sp. and S. terebellae Ray 1930 demonstrate niche partitioning in marine gregarine parasites (apicomplexa)', *Protist*, 165: 493-511.

Waldeland, H., and J. K. Frenkel. 1983. 'Live and killed vaccines against toxoplasmosis in mice', *J Parasitol*, 69: 60-5.

Walker, R. A., D. J. Ferguson, C. M. Miller, and N. C. Smith. 2013. 'Sex and Eimeria: a molecular perspective', *Parasitology*, 140: 1701-17.

Walliker, D. 1991. 'Malaria Parasites - Randomly Interbreeding or Clonal Populations', *Parasitology Today*, 7: 232-35.

Walzer, K. A., and J. P. Boyle. 2012. 'A single chromosome unexpectedly links highly divergent isolates of Toxoplasma gondii', *MBio*, 3.

Wang, J. L., T. T. Li, H. M. Elsheikha, K. Chen, W. N. Zhu, D. M. Yue, X. Q. Zhu, and S. Y. Huang. 2017. 'Functional Characterization of Rhoptry Kinome in the Virulent Toxoplasma gondii RH Strain', *Front Microbiol*, 8: 84.

Wang, Z. T., S. K. Verma, J. P. Dubey, and L. D. Sibley. 2017. 'The aromatic amino acid hydroxylase genes AAH1 and AAH2 in Toxoplasma gondii contribute to transmission in the cat', *PLoS Pathog*, 13: e1006272.

Wardal, E., K. Markowska, D. Zabicka, M. Wroblewska, M. Giemza, E. Mik, H. Polowniak-Pracka, A. Wozniak, W. Hryniewicz, and E. Sadowy. 2014. 'Molecular analysis of vanA outbreak of Enterococcus faecium in two Warsaw hospitals: the importance of mobile genetic elements', *Biomed Res Int*, 2014: 575367.

Wasmuth, J. D., V. Pszenny, S. Haile, E. M. Jansen, A. T. Gast, A. Sher, J. P. Boyle, M. J. Boulanger, J. Parkinson, and M. E. Grigg. 2012. 'Integrated bioinformatic and targeted deletion analyses of the SRS gene superfamily identify SRS29C as a negative regulator of Toxoplasma virulence', *MBio*, 3.

Watts, E., Y. Zhao, A. Dhara, B. Eller, A. Patwardhan, and A. P. Sinai. 2015. 'Novel Approaches Reveal that Toxoplasma gondii Bradyzoites within Tissue Cysts Are Dynamic and Replicating Entities In Vivo', *MBio*, 6.

Wendte, J. M., A. K. Gibson, and M. E. Grigg. 2011. 'Population genetics of Toxoplasma gondii: new perspectives from parasite genotypes in wildlife', *Vet Parasitol*, 182: 96-111.

Wendte, J. M., M. A. Miller, D. M. Lambourn, S. L. Magargal, D. A. Jessup, and M. E. Grigg. 2010. 'Self-mating in the definitive host potentiates clonal outbreaks of the apicomplexan parasites Sarcocystis neurona and Toxoplasma gondii', *PLoS Genet*, 6: e1001261.

Wendte, J. M., M. A. Miller, A. K. Nandra, S. M. Peat, P. R. Crosbie, P. A. Conrad, and M. E. Grigg. 2010. 'Limited genetic diversity among Sarcocystis neurona strains infecting southern sea otters precludes distinction between marine and terrestrial isolates', *Vet Parasitol*, 169: 37-44.

West, S. A., T. G. Smith, and A. F. Read. 2000. 'Sex allocation and population structure in apicomplexan (protozoa) parasites', *Proc Biol Sci*, 267: 257-63.

Wootton, J. C., X. Feng, M. T. Ferdig, R. A. Cooper, J. Mu, D. I. Baruch, A. J. Magill, and X. Z. Su. 2002. 'Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum', *Nature*, 418: 320-3.

Woznica, A., A. M. Cantley, C. Beemelmanns, E. Freinkman, J. Clardy, and N. King. 2016. 'Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates', *Proc Natl Acad Sci U S A*, 113: 7894-9.

Woznica, A., J. P. Gerdt, R. E. Hulett, J. Clardy, and N. King. 2017. 'Mating in the Closest Living Relatives of Animals Is Induced by a Bacterial Chondroitinase', *Cell*, 170: 1175-83 e11.

Wyrosdick, H. M., and J. J. Schaefer. 2015. 'Toxoplasma gondii: history and diagnostic test development', *Anim Health Res Rev*, 16: 150-62.

Xiao, J., and R. H. Yolken. 2015. 'Strain hypothesis of Toxoplasma gondii infection on the outcome of human diseases', *Acta Physiol (Oxf)*.

Yamamoto, M., J. S. Ma, C. Mueller, N. Kamiyama, H. Saiga, E. Kubo, T. Kimura, T. Okamoto, M. Okuyama, H. Kayama, K. Nagamune, S. Takashima, Y. Matsuura, D. Soldati-Favre, and K. Takeda. 2011. 'ATF6beta is a host cellular target of the Toxoplasma gondii virulence factor ROP18', *Journal of Experimental Medicine*, 208: 1533-46.

Yamamoto, M., M. Okuyama, J. S. Ma, T. Kimura, N. Kamiyama, H. Saiga, J. Ohshima, M. Sasai, H. Kayama, T. Okamoto, D. C. Huang, D. Soldati-Favre, K. Horie, J. Takeda, and K. Takeda. 2012. 'A cluster of interferon-gamma-inducible p65 GTPases plays a critical role in host defense against Toxoplasma gondii', *Immunity*, 37: 302-13.

Yan, C., L. J. Liang, K. Y. Zheng, and X. Q. Zhu. 2016. 'Impact of environmental factors on the emergence, transmission and distribution of Toxoplasma gondii', *Parasit Vectors*, 9: 137.

Yang, C. S., J. M. Yuk, Y. H. Lee, and E. K. Jo. 2015. 'Toxoplasma gondii GRA7-Induced TRAF6 Activation Contributes to Host Protective Immunity', *Infection and Immunity*, 84: 339-50.

Yang, N. H., A. Farrell, W. Niedelman, M. Melo, D. N. Lu, L. Julien, G. T. Marth, M. J. Gubbels, and J. P. J. Saeij. 2013. 'Genetic basis for phenotypic differences between different Toxoplasma gondii type I strains', *Bmc Genomics*, 14.

Yarovinsky, F., D. Zhang, J. F. Andersen, G. L. Bannenberg, C. N. Serhan, M. S. Hayden, S. Hieny, F. S. Sutterwala, R. A. Flavell, S. Ghosh, and A. Sher. 2005. 'TLR11 activation of dendritic cells by a protozoan profilin-like protein', *Science*, 308: 1626-9.

Yera, H., D. Ajzenberg, F. Lesle, D. Eyrolle-Guignot, M. Besnard, A. Baud, A. Georges, J. Dupouy-Camet, and M. L. Darde. 2014. 'New description of Toxoplasma gondii genotypes from French Polynesia', *Acta Trop*, 134: 10-2.

Yin, T., D. Cook, and M. Lawrence. 2012. 'ggbio: an R package for extending the grammar of graphics for genomic data', *Genome Biol*, 13: R77.

Zhang, J., A. Khan, A. Kennard, M. E. Grigg, and J. Parkinson. 2017. 'PopNet: A Markov Clustering Approach to Study Population Genetic Structure', *Mol Biol Evol*.

Zhang, K., G. Lin, Y. Han, and J. Li. 2016. 'Serological diagnosis of toxoplasmosis and standardization', *Clin Chim Acta*, 461: 83-9.
Zhang, N. Z., M. Wang, Y. Xu, E. Petersen, and X. Q. Zhu. 2015. 'Recent advances in developing vaccines against Toxoplasma gondii: an update', *Expert Rev Vaccines*, 14: 1609-21.
Zhang, N. Z., Y. Xu, S. Y. Huang, D. H. Zhou, R. A. Wang, and X. Q. Zhu. 2014. 'Sequence variation in Toxoplasma gondii rop17 gene among strains from different hosts and geographical locations', *ScientificWorldJournal*, 2014: 349325.
Zhao, Y., and G. S. Yap. 2014. 'Toxoplasma's arms race with the host interferon response: a menage a trois of ROPs', *Cell Host Microbe*, 15: 517-8.
Zhou, P., H. Zhang, R. Q. Lin, D. L. Zhang, H. Q. Song, C. Su, and X. Q. Zhu. 2009. 'Genetic characterization of Toxoplasma gondii isolates from China', *Parasitol Int*, 58: 193-5.

Curriculum Vitae

Andrea Kennard

Cell: (720) 771-7220
andrea.kennard@jhu.edu

EDUCATION

2011-2017 PhD Candidate, Cell, Molecular, Developmental Biology, and Biophysics, **Johns Hopkins University**, Baltimore, MD
Advisor: Dr. Mike E. Grigg

2009 B.A., Molecular, Cellular, and Developmental Biology, **The University of Colorado**, Boulder, CO

RESEARCH EXPERIENCE

2011-Present *Ph.D. Candidate*, Program in Cell, Molecular, Developmental Biology, and Biophysics, Johns Hopkins University-National Institutes of Health. The Role of Sexual Recombination in the Evolution of the Protozoan Parasite, *Toxoplasma gondii*.

2009-2011 *Research Assistant*, University of Maryland Baltimore, Microbial Pathogenesis Department, Baltimore, MD. Cloning and characterization of polymorphic outer membrane protein expression and clinical sample isolation in *Chlamydia.*

Fall 2007 *Student Researcher*, University of Colorado Python Project Laboratory, Boulder, CO. Characterization of the python UCP-2 gene within cardiac hypertrophy.

FELLOWSHIPS AND AWARDS

2011- Present NIH Intramural Research Training Award (IRTA) Fellowship
2009 Phi Beta Kappa Honor Society Acceptance
2006-2008 Dean's List University of Colorado, Boulder

PUBLICATIONS

1. **Kennard, A.**, Khan, A., Miller, M.A., James, E.R., Miller, N., Roos, D.R., Keeling, P., Conrad, P.A., Grigg, M.E. (2018) Virulence Shift in a Sexual Clade of Type X *Toxoplasma* Infecting Southern Sea Otters. *In Progress.*
2. Zhang, J., Khan, A., **Kennard A**., Grigg, M.E., Parkinson, J. PopNet: A Markov Clustering Approach to Study Population Structure. (2017) Molecular Biology and Evolution 34,1799-1811.

3. Chudnovskiy, A., A. Mortha, V. Kana, **A. Kennard**, A. Rahman, R. Remark, I. Mogno, R. Ng, S. Gnjatic, A. Solovyov, B. Greenbaum, J. Clemente, J. Faith, Y. Belkaid, M.E. Grigg, M. Merad. Host-protozoan interactions shape mucosal immunity and protect from bacterial infections through activation of the inflammasome. (2016) Cell, Vol 167, Issue 2, 444-456**.**

4. Pszenny, V., Ehrenman, K., Romano, J.D., **Kennard, A.**, Schultz, A., Roos, D.S., Grigg, M.E., Carruthers, V.B., Coppens, I. A Lipolytic Lecithin: Cholesterol Acyltransferase Secreted by Toxoplasma Facilitates Parasite Replication and Egress. (2016) J Biol Chem 291, 3725-3746.

MEETINGS AND PRESENTATIONS

2017 CIFAR Integrated Microbial Diversity Program Meeting, Whistler, BC, Canada, June 7. Poster Presentation, "Cryptic Virulence Among Type X Toxoplasma: A Recombinant Clade Infecting North American Wildlife*."*

2016 NIH Graduate Student Research Symposium, Bethesda, MD, January 12. Poster Presentation, "Cryptic Virulence Among Type X Toxoplasma: A Recombinant Clade Infecting North American Wildlife*."*

2015 International Congress on Toxoplasmosis & Toxoplasma gondii Biology, Gettysburg, PA, June 19. Poster Presentation, "Cryptic Virulence Among Type X Toxoplasma: A Recombinant Clade Infecting North American Wildlife*."*

2015 NIH Graduate Student Research Symposium, Bethesda, MD, January 13. Poster Presentation, "Sex Happens: Evidence of Recombination within a Clade of *Toxoplasma gondii."*

2014 NIH Graduate Student Research Symposium, Bethesda, MD, January 14. Poster Presentation, "Genetic Basis for Host Range Expansion of the Cosmopolitan Parasite *Toxoplasma gondii."*

2013 NIH Graduate Student Research Symposium, Bethesda, MD, January 15. Poster Presentation, "Genetic Basis for Host Range Expansion of the Cosmopolitan Parasite *Toxoplasma gondii."*

2011 Chlamydia Basic Research Society Conference, Redondo Beach, CA, March 18. Attendee, Poster Presented, "Expression of the polymorphic membrane protein gene family of Chlamydia caviae under conditions of penicillin-induced persistence."

TEACHING EXPERIENCE

Spring 2015 Lecturer, Research Tools for Studying Disease (~30 students)
Fall 2014 Scientists Teaching Science Lecturer Course
Spring 2012 Teaching Assistant, Cellular Biology Laboratory (~20 students)
Fall 2011 Teaching Assistant, Biochemistry Laboratory (~20 students)

MENTORING

Summer 2017 Student Mentor, Bnos Yisroel High School Bridge Program – Sara Verschleisser – Epstein-Barr Virus Deregulation of MYC and BCL2L11 Contributes to the Genesis of Burkitt's Lymphoma

LEADERSHIP/COMMITTEES

2015 Member, NIH Graduate Partnerships Program Retreat Committee

PROFESSIONAL ACTIVITIES

2017 Lead Poster Judge, NIH Post baccalaureate Poster Day, May 4
2016 Lead Poster Judge, NIH Post baccalaureate Poster Day, April 20
2015 Poster Judge, NIH Post baccalaureate Poster Day, April 30
2013-2014 STEM Learning Studios Teaching Outreach with National Commission on Teaching and America's Future (NCTAF) Scientific Partner

PROFESSIONAL DEVELOPMENT

2017 Integrated Microbial Biodiversity Trainee Workshop
2016 NIH OITE Translational Science Training Program Bootcamp
Fall 2014 Scientists Teaching Science Lecturer Course
Spring 2014 Introduction to Python – NIH
Fall 2013 Practical Bioinformatics Class – NIH
Fall 2013 Making Figures in R